

论 信 息

The Nature of Information

杨学山 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书在分析各类信息的基础上,重新定义了信息,归纳了信息相异于物质、能量、生命的特殊属性,剖析了信息的各种运动形态,总结了基于含义的信息结构,讨论了信息的增长过程,阐述了信息独特的发展逻辑,勾画了信息空间及其与物理空间、生命空间的关系及由物理空间、生命空间和信息空间构成的宇宙发展新图像。

本书用简明易懂的语言和逻辑回答了信息是什么,诠释了信息发生发展的规律,适合广泛人群阅读。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

论信息 / 杨学山著. —北京:电子工业出版社, 2016.4

ISBN 978-7-121-28349-9

I. ①论... II. ①杨... III. ①信息论 IV. ①G201

中国版本图书馆 CIP 数据核字(2016)第 052054 号

责任编辑:董亚峰

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:720×1 000 1/16 印张:21 字数:322 千字

版 次:2016 年 4 月第 1 版

印 次:2016 年 4 月第 1 次印刷

定 价:88.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 zltts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

前 言

信息，平凡而又神秘。说其平凡，是因为几乎所有人都在用信息这个词；说其神秘，是因为居然到现在还没有一个具有共识的定义，还没有形成体系化的、揭示其发生发展规律的知识体系。然而，人类社会正在走向信息社会。依据物理学发现的材料和能源技术推动了工业革命数百年的发展，同样的科学理论能否解释信息发生发展的规律并推动信息社会可持续发展呢？这是本书的起点。

本书的目的就是明确信息定义，揭示信息发生发展的规律。本书共六章，第 1 章给出了信息的定义和属性，第 2 章归纳信息的各种运动形态，第 3 章介绍了基于含义的信息结构，第 4 章讨论信息的增长过程及规律，第 5 章阐述信息的独特的不同于物质和生命的发展逻辑，第 6 章介绍信息空间及其与物理空间、生命空间的关系，勾画了由物理空间、生命空间和信息空间构成的宇宙的发展图像。

信息是什么，让太多的研究者困惑了太长时间。不是因为认知的复杂性，而是因为信息与物质、能量和生命全然相异的特殊性，破解这个特殊性，信息这种宇宙中唯一的特殊存在就跃然纸上。第 1 章定义了信息，并分析了信息的这种特殊性。本书定义的信息就是所有客观存在的含义，这里，物理空间、生命空间和信息空间都是客观存在的。如此定义的信息没有质量、没有能量、没有固定形态，如此定义的信息必须依托载体与/或外壳存在，如此定

论信息

义的信息可以在不同的形态间转换，如此定义的信息借助不同的载体与/或外壳逐步走向自主发展，最终成为独立的自我发展的与物理空间、生命空间并列的信息空间。如此定义的信息以自在态、自有态和记录态三种模式存在，可以解释几乎所有今天在使用的信息概念。信息由载体、外壳、含义三要素构成；具有三种截然不同的存在形态，含义可以在这三种形态中不断转换；含义本身不受时空和物理定律、生命定律的约束，跨越时空和物质、能量、生命，实现非连续的连续进化，这是信息区别于物质、能量、生命的三个最重要的特殊性，是信息的一系列特殊属性产生的根源，也是认识信息、理解信息的入门钥匙。

区分信息的形与义是认识信息及信息发生发展规律的关键。信息的载体和外壳是形，含义是义。承载信息的光、声、电、纸张、核苷酸、神经元是形，文字、语言、图表、概念是外壳，核苷酸、神经元、非生物主体存储处理的外壳既是载体又是外壳。载体和外壳都是形，而可以为认知主体利用的含义才是义。载体及与载体结合在一起的外壳运动或遵循物理空间定律，或遵循生命空间定律，而含义的运动和发展不遵循这两类定律，因此，无论是信息的运动、结构、增长、逻辑和空间，都需要区分这两者，作为分析和讨论的基础。

信息的运动既有以形为主的，也有以义为主的，还有形与义同时运动的，第 2 章从五个维度讨论了信息的运动，对于基于含义的信息增长和演进这两类本质上属于信息运动内容的讨论留给了后续章节。信息运动从三态转换开始，这是信息发生发展的基础，从自在态信息到遗传信息、从遗传信息到认知信息、从认知信息到记录信息，这几个关键的转换环节在几十亿年时间内逐一突破，为最终形成独立、自主发展的信息空间奠定了基础。而信息形与义的移动、处理则揭示了不同形态信息移动和处理的同异之处，载体和外

壳遵循各自所属形态的规律，而含义的移动和处理都是朝着特定的目的，遗传、认知、记录信息的处理具有相同的模式，今天计算机信息处理能力与遗传和认知信息的处理相比，在形的处理上超出很多，但在义的处理上还有很大的差距。第 2 章的最后两节只讨论信息载体与/或外壳的积累和演进。其中记录态信息的积累及与记录态信息发展连接在一起的符号、概念体系演进是信息空间演进过程中的基础性、标志性环节。

在不同的信息形态和不同发展阶段，保持信息自身持续增长的基础是信息结构。没有结构的自在态信息无论数量多少、形态如何均是一盘散沙，不能发挥作用，更不能实现有效的基于含义的增长。信息结构是理解信息、把握信息发生和发展规律的钥匙。第 3 章的目的是介绍信息结构的基本概念及主要特征。信息结构的基础是信息单元，如同生命空间的细胞和物理空间的分子，既由更基本的组成部分构成，又具有独立的功能。信息结构有两类，隐性信息结构和显性信息结构。隐性信息结构只能为拥有该信息集合的主体所利用，显性信息结构可以为多个主体利用。当自有信息产生之日起，隐性信息结构就存在，并不断发展，记录信息产生之日起，显性信息结构就开始产生，记录信息本身就是带有信息结构的。随着记录信息的不断增长和信息处理技术的不断发展，显性信息结构在逐步显示其特征和作用，从互联网大数据到不同类型的智能系统，显性信息结构都是其主要的构成因素。然而，所有这一切都没有被称为显性信息结构，而是被称为索引、目录、知识表示等。正如遗传信息编码和大脑认知信息存储和利用结构没有破解之前，没有认识到隐性信息结构一样。正如隐形信息结构才使遗传信息和认知信息成为生命和认知持续发展的基础，显性信息结构是记录信息自我发展的基础。本章也介绍了信息结构的一般特征和基本要求。

第4章定义了信息的增长并划分了信息增长的六个发展阶段，重点分析了信息结构增长的一般模式和路径。信息增长既是信息的量和质的增长过程，更是信息空间功能逐步完善直至自我发展的过程。第一阶段到第三阶段是漫长的借态演进过程，使信息空间从没有生机的无限沉寂状态，跨越到生成不依赖物理和生命空间独立存在的记录态信息。这个过程以物理空间和生命空间的定律为基础，完成了信息空间进化最为关键的两步：遗传基因的进化并通过遗传基因，使智人的进化达到了10万年前高度完备的认知功能，为成熟的体系化的语言和文字、符号和概念体系的诞生奠定了不可逆的雄厚基础。相比于第一到第三阶段的40多亿年时间（仅从地球看），第四、第五两个阶段只有不足一万年时间，但实现了信息空间功能增长的另一个飞跃，认知能力和基于认知能力的记录信息增长，从串行进化发展到大规模并行进化。此前关于信息的演进都是基于个体，数量众多的个体按照生命进化规律发展，值得特别关注的是，关于认知能力的生命进化似乎就以智人为终点。极其大量的生命个体经历数十亿年时间，完成了认知能力功能的进化，这是生命规律的约束。一旦摆脱了生命进化规律，记录信息诞生之后，生命规律和社会发展规律同时发挥作用，使信息空间在短短几千年时间内发生了三个重大的变化，一是记录信息载体与/或外壳的处理工具不断进化，远远超出了人的处理能力；二是借助记录信息处理工具，人类认知能力的进化并行发展规模越来越大，从开始的数十、数百到现在的几十亿在一个共享的认知信息平台上发展；三是正因为快速地并行认知进化，很快就耗尽了生命空间赋予人类的认知能力，人类认知能力极限已经触及，半个世纪以来人们称之为“科学的沉寂”现象，在很大程度上是人类认知能力的极限，逐步完备的显性信息结构一定程度上可以缓解这一进程，但极限的到来是必然的。显性信息结构的诞生和逐步完备、信息处理技术体系的发展和人类认知能力的极限三个因素

促使信息发展进入第六阶段，尽管这个阶段还没有到来，但轮廓已经显现，就是一个独立于物理和生命空间，自我发展完善的信息空间的产生，并与物理和生命空间一起成为三足鼎立的宇宙发展新格局。

理解信息增长和信息空间的核心是显性信息结构。显性信息结构的基础是隐性信息结构，显性信息结构只是将隐性信息结构外化，成为不同主体可以共享，并能在一定条件下重现客观存在的信息空间基础构成要素。一个显性信息结构对应于一个客观存在，包括这个客观存在的属性和功能。第 4 章第五节定义了显性信息结构完备的充要条件，满足这个条件的显性信息结构就使信息空间得到了自我生长与完善的基础。当第六阶段信息空间认知主体的能力不断提升，显性信息结构完备的理论、方法和工具不断完善，越来越多的隐性信息结构将转换为显性信息结构，物理空间和生命空间未知的规则将逐一被认知，信息发展第六阶段将进入成熟期。

信息运动的规律及规律蕴含的逻辑是什么，这是第 5 章的主要内容。信息逻辑的讨论有感知、连接、计算和进化这四个含义，以运动为主要领域展开。感知与转换是一个将一个（或多个）主体的外部信息转换为内部信息，成为隐性或显性信息结构的一个组成部分的信息运动过程。这个运动过程按感知的主体和载体的不同，分别遵循物理或生命规律，但决定感知什么、取舍什么、含义如何组织，进入主体什么信息集合或结构则遵循信息的规律。连接是信息运动最重要的组成部分，连接的实现遵循物理或生命定律，连接的目的和结果基于信息规律。按照信息空间的划分，信息的连接分成内连接和外连接，按照连接建立的性质，分为实连接、虚连接和任意连接三类。这三类连接构成了基于含义的信息空间连接场，根据一定的规则实现不同目的和节点间的连接。信息的处理与信息的连接一样无时无刻不在。对于含义的处理与外壳或载体的处理性质截然不同，因为客观存在本身是由一个个基本

构件组成的，每个基本构件都有其特定的处理功能，以信息结构为基础的处理也应该是从最基本的构建开始。因此，处理的逻辑由一组微处理、短逻辑起始，以内计算、泛包容、多中心为规则构成信息空间的处理模式。以信息结构内成的功能系列为基础，以多中心计算模式，可以形成大任务、泛分解的处理模式。通过感知转换功能和任意连接、内计算、泛包容的规则，使以显性信息结构为基础的信息空间的计算，成为一个以内部完善为主的计算体系，为信息空间的自我完善奠定了动力基础。关于集合计算和多维计算的规则提供了信息空间，另一类计算对象新的处理能力。按照物理和生命空间的思维逻辑看信息空间的增长，被割裂为一个个不同的没有任何逻辑关系的片段，从而成为物理或生命空间的一个附属现象。换一个思维模式，从信息含义在不同性质的空间中，在物理和生命定律的作用下，形成自身的发展路径和规律，可以看到这种非连续的连续进化的必然性和规律性，第 5 章的第六节解释了这一特殊的生长过程。

第 6 章是对信息空间的一个总结性分析章节，主要是三个部分，一是信息空间的构成及其存在、消亡与扩展，二是信息空间与认知能力发展的关系，三是信息空间与物理和生命空间的关系。信息空间的发展还没有完成一个循环，总结其规律还为时过早，但迄今为止的发展，至少可以得出三条规律性的总结：一是在一个独立信息空间中，转换主体的数量、延续的时长、协同平台的质量是该信息空间规模增长和进化速度决定了该空间增长的速度，称之为延续协同定律。二是信息空间的发展呈台阶式跃迁，每一个台阶呈现长期渐进积累后跃迁到一个新台阶规律，称之为积累跃迁定律。三是信息空间构成的非组合规律，一对一地反映相应的客观存在，一对一地在隐形或显性信息结构中体现，一对一地复制、处理。这正是信息空间的构成特征，也是信息空间的生长特征。

信息发展六个阶段是信息空间形成和完善的过程，理解信息空间，必须摆脱物质世界和生命世界延续性的定式，这一种不同于物质和生命的规律主导着信息空间的延续。物理空间定律必然产生自在态信息并与生物体发生交互，生命空间定律必然导致遗传基因的产生进化并达到生物体认知功能的极限，却诞生了独立于物理和生命空间，不遵循物理和生命定律而发生发展的信息空间。

第6章在6.2.5节中提出了一个重要的假设，即如果信息是物理空间的组成部分，是物质基本粒子的构成部分，则将在信息空间产生第二类自在态信息，改变物理、生命、信息空间相互的关系。物理学既没有证明这种状态的存在，也没有证明这种可能不存在，所以只能是一种假设。如果这个假设成立，信息是物理空间物质的组成部分，而不仅是反映客观存在的。分子生物学的研究成果已经证明，信息是生命空间的必要组成部分。信息由载体、外壳和含义构成，有三种存在形态，在信息进化过程中，生命空间是其关键阶段的必要条件，信息空间依赖物理和生命空间而存在。

三个空间相互依存，相互影响，形成新的宇宙观。从宇宙演进的维度看，物理空间的逻辑是能量、质量的守恒、力的作用和统一，在物理空间多变的现象中寻找不变的规律；生命空间的逻辑是在物理空间逻辑的约束下，在不变中寻找可变的生存和发展；信息空间的逻辑是在物理空间和生命空间逻辑的约束下，寻找改变物理空间和生命空间运行轨迹的路径。这里有两个关键的环节，一是是否存在通过信息影响物质的微观构成和宏观进程的可能，二是物理、生命、信息三个空间谁能主导认知能力。对于第一个环节，今天的物理学还不能给出答案，但直接否定是不成立的。对于第二个环节，物理空间不能主导认知，主流的意见认为人类是认知的主体，至少人类是主导认知能力的主体。但是，认知所要揭示的客观规律的复杂

论信息

性与认知能力的提升同步增加、人类生理极限的矛盾和信息空间自主认知能力的加速成长，如果找不到恰当的解决办法，非生物体超越人的认知能力的时间已经不会很长。对于这一问题的进一步论述，我将在 2017 年出版的《智能原理》一书中展开。

本书试图将信息这个客观存在清晰定义，并在此基础上揭示其发生发展的规律，并以信息空间的发展及其与物理、生命空间的关系，重新诠释宇宙的发展规律。30 多年的研究历程，无数次的推翻和肯定，最终形成了这一书稿。它还不成熟，还有很多可以推敲和完善的地方，甚至可能存在错误的结论，但我还是把它出版了，目的就是得到读者的批评，以完善这一理论体系。

本书的前言和目录由美国长岛大学储荷婷教授翻译成英文，这是一个高难度的创作过程，谨此表示感谢。

感谢电子工业出版社的刘九如、秦绪军、董亚峰在本书出版过程中付出的辛勤劳动，是他们的智慧和工作，提高了本书的质量和进度。

Preface

Information, a term that is ordinary yet mysterious. The term is ordinary because almost everyone uses it. It is mysterious in that there is still no agreed-upon definition for information. Nor is there a knowledge framework that systematically illustrates the emergence and development of information. The human society is, nevertheless, moving towards the information society. Physics-based discoveries in materials and energy technology provided the impetus to the Industrial Revolution for several hundred years. Will the same scientific theory be able to explain the laws that govern the emergence and development of information and promote sustainable development of the information society? This is the start point of the present book.

The purpose of this book is to clearly define information and to uncover the laws governing the emergence and development of information. This book consists of six chapters. Chapter 1 defines information and describes its attributes. Chapter 2 summarizes various motions of information. Chapter 3 depicts meaning-based information structure. Chapter 4 discusses the process and pattern of information growth. Chapter 5 elaborates the unique development logic of information that is different from matter and life. Chapter 6 covers information space and its relations with the physical space and life space, portraying the development of the universe that is composed of the physical space, life space and information space.

What is information? This question has bewildered so many researchers for too long, which is not due to the complexity of cognition but caused by the

particularity of information that is entirely different from matter, energy and life. Information apparently becomes the only unique existence in the universe after its particularity is deciphered. Chapter 1 defines information and analyzes its particularity. Information in this book is defined as “the connotations of all objective existence”. The physical space, life space and information space are all objective existence. Information so defined has no mass, energy and fixed form. It must rely on some medium (e.g., paper) and/or shell (e.g., language) to exist. It can transform among different states (i.e., self-existent, self-containing, recorded) of information. Information so defined would develop autonomously via different media and/or shells, and ultimately become an independent information space that is parallel to the physical space and life space. It exists in the forms of self-existent, self-containing, and recorded, which can explain almost all the information concepts in use today. Information comprises three key components (i.e., medium, shell, meaning). It exists in three completely different states (i.e., self-existent, self-containing, recorded) while its meaning can constantly change as it transforms from one form to another; The meaning of information itself is not constrained by time, space, physics laws and life laws as well as transcends space, time, matter, energy and life to make the disrupted continuous development. All the above constitute the three most importance characteristics that separate information from matter, energy and life. These three characteristics become the source for a series of special attributes of information. They also serve as the entry key to learning about and understanding information.

Distinguishing the form and meaning of information is the linchpin for understanding information and the principles determining its emergence and development. “Form” refers to the medium and shell of information while “meaning” concerns about the connotation of information. Form includes light, sound, electricity, paper, nucleotides and neurons that function as the medium of information whereas words, languages, graphics and notions are shells. The shell of nucleotides, neurons and other non-biological subject storage processing also

serves as the medium. As indicated earlier, form refers to both the medium and shell while meaning is the connotation that the cognitive subject can utilize. Media and the motion of shells that are combined with media follow either the physical space laws or the life space laws. In contrast, the motion and development of meaning do not follow these two types of laws. Therefore, we need to distinguish between form and meaning no matter whether we consider information motion, structure, growth, logic or space. This constitutes the basis for our analysis and discussion of information.

The motion of information could mainly rely on form or meaning. It is also possible that form and meaning involve in information motion simultaneously. Chapter 2 discusses the five aspects of information motion. As for meaning-based information growth and development – the two topics that fit in the domain of information motion by nature, we will explore them in subsequent chapters. Information motion starts with the transformation among the three states (i.e., self-existent, self-containing, recorded) and four categories (i.e., self-existent, hereditary, cognitive, recorded): from self-existent information to hereditary information, from hereditary information to cognitive information, from cognitive information to recorded information. These three crucial transformations are accomplished one by one in billions of years, laying the foundation for ultimately forming the information space that develops independently and autonomously. On the other hand, the move and processing of information form and meaning illustrate the differences and similarities between that of information in various states (i.e., self-existent, self-containing, recorded). Media and shells follow the patterns of their respective states whereas the move and processing of meaning aim toward a particular objective. Processing of hereditary, cognitive and recorded information shares a similar model. Compared with hereditary and cognitive information processing, computer information processing capability today is much higher in terms of the form but remains lagging behind by a big margin as far as the meaning is concerned. The last two sections of Chapter 2 are devoted to the

discussion on the accumulation and development of information media and/or shells. The accumulation of recorded information along with symbols and conceptual framework development that associate with recorded information is the basic and iconic steps in the advancement of information space.

Information structure is the foundation for maintaining sustained growth of information in different states and at different development stages. Without any structure, information of the self-existent state is just a mess no matter how large the quantity is. It also cannot play any role in the information space, let alone achieve effective, meaning-based growth. Information structure is the key to understanding information and finding out the laws of information emergence and development. The purpose of Chapter 3 is to present basic concepts and main characteristics of information structure. The foundation of information structure consists of information units. Like cells in the life space and molecules in the physical space, information units are composed of more basic elements and yet have independent functionalities. There are two types of information structure: latent and manifest. The latent information structure can only be utilized by the subject who possesses the information set while the manifest information structure can be employed by multiple subjects. The latent information structure has been in existence and kept growing ever since the emergence of self-containing information. The manifest information structure has started forming since the emergence of the recorded information. In fact, recorded information itself holds information structure. The characteristics and functionalities of the manifest information structure become more and more apparent as recorded information continues growing and information processing technology continues developing. From the Internet, Big Data to intelligent systems of different kinds, they all have the manifest information structure as their major component. Yet all these are not regarded as manifest information structures. Rather, they are referred to as indexes, catalogs, knowledge representation, and the like. The latent information structure is not recognized before the hereditary information coding and brain cognitive

information storage and usage mechanism are understood. On the other hand, it is the latent information structure that makes hereditary information and cognitive information become the foundation for the sustained development of life and cognition. Likewise, the manifest information structure is the foundation for the self-development of recorded information.

Chapter 4 defines information growth and then divides it into six development stages with analyses focusing on the general model and approach of information structure growth. Information growth represents the process not only for information to grow in both quantity and quality but also for the function of information space to gradually improve till it reaches the position of self-development. The first stage to the third stage of information growth is a long development process which enables the information space to advance from the lifeless, motionless and soundless state to the recorded state of information which exists independently and whose generation does not rely on the physical and life spaces. On the basis of the physical space and life space laws, this process completes the two most crucial steps in information space development: 1) the evolution of heredity genes, and 2) The Homo sapiens, through the heredity genes, evolved to attain the cognitive ability that was considered highly complete one hundred thousand years ago. This lays an irreversible and solid foundation for the emergence of matured and systematized languages, words, symbols and conceptual systems. Compared with the more than four billion years between the first stage and the third stage (only from the perspective of the Earth), there are less than ten thousand years between the fourth stage and fifth stage. But another leap is made in enriching information space function. That is, cognitive capability and information growth based on cognitive capability progress from serial development to large-scale parallel development. All the information development before then is individual-based. Numerous individuals develop in accordance with the law of life evolution. What deserves special attention is that the destination for the evolution of life in terms of cognitive ability appears to be the Homo sapiens. An extremely large

number of individual lives complete the evolution of their cognitive abilities over the course of billions of years, which is bound by the law of life evolution. Once beyond the law of life evolution and after the emergence of recorded information, the law of life and the law of social development take effect simultaneously which leads to three important changes in information space during the short period of thousands of years (as opposed to the billions of years discussed earlier): 1) The continuous improvement of processing tools of recorded information media and shells surpasses the processing capacity of the mankind; 2) With the help of the recorded information processing tools, the evolution and parallel development of human cognitive ability is on a scale that becomes larger and larger. It starts with tens, hundreds of individuals till the development of a cognitive information platform shared by billions. 3) It is because of the rapid development of parallel cognition that quickly exhausts the cognitive ability with which the mankind are endowed from the life space. As a result, the mankind are reaching their limit of cognitive ability. The so-called “science stagnation and stillness” phenomenon occurring in the past half a century to a large extent is due to this limit of human cognitive ability. Improving the manifest information structure in a certain degree can slow down such a process. But it is inevitable for the mankind to reach the limit of cognitive ability one day. The emergence and gradual completion of the manifest information structure, the development of information processing technology system and the limit of human cognitive capability constitute the three factors that together push information development to its sixth stage. Although this stage has not yet come, its contour has been arising: The emergence of an information space which is independent of the physical and life spaces and capable of self-development and improvement. This information space, together with the physical space and life space, become the three pillars for supporting the development of the universe.

The core for understanding information growth and information space is the manifest information structure. Yet the latent information structure forms the basis

of the manifest information structure. In other words, the manifest information structure externalizes the latent information structure and can be shared by different subjects. In addition, it is capable of representing the basic key components of information space for all objective existence under certain conditions. One manifest information structure corresponds to one objective existence, including its attributes and functions. Chapter 5, Section 5 defines the prerequisites and requirements for the completeness of manifest information structure. The manifest information structure that meets those prerequisites and requirements enables the information space to have a foundation for its self-development and improvement. When the information space reaches its sixth stage, the abilities of cognitive subjects keep increasing. Theories, methods and tools for completing manifest information structures also continue to improve. Likewise, more and more latent information structures are transformed into manifest ones. Unknown patterns and principles for the physical space and life space are discovered one by one. This signifies that the sixth stage of information development will enter into maturity.

What is the pattern for information motion and the logic implied by the pattern? These are the major topics Chapter 5 intends to address. Discussion on information logic is to be conducted in four key areas: conception, connection, computation and development. Perception and transformation refer to the process in which the external information of one (or more) subject is transformed to internal information and thus becomes one component of the latent or manifest information structure in information motion. This motion process follows respective physical or life laws depending on the subjects of perception and different media. Nevertheless, it follows the information law when determining what to perceive, what to select, how to organize contents, and which information set or structure of the subject to enter. Connection is the most important component in information motion. Connections are made by following the physical and life laws while the purpose and result of connection is based on the information law.

Information connections are categorized as internal connections and external connections in accordance with the division of information space. Information connections can also be grouped into three kinds: real, virtual and arbitrary in accordance with the nature of the connection is established. These three kinds of connections constitute the meaning-based information space connection field which, according to certain specifications, makes connections among different knots for different purposes. Like information connections, information processing is ubiquitous and present at any time. But the nature of information processing is entirely different in dealing with meanings and in handling media or shells. This is because all objective existence is made of individual basic components and every basic component has its own specific processing function. Processing on the basis of information structure should start with the most basic components. Therefore, processing logic starts with a series of micro-processing and short logic, forming a processing model of information space that comprises the rules of internal computation, universal inclusion and multi-center computation. The processing model can be constructed for handling big tasks and universal decompositions by building on a series of functions generated within the information structure and by taking the multi-center computation model. Through the transformation function of perception as well as the specifications for arbitrary connections, internal computation and universal inclusion, information space computation based on the manifest information structure would become a computation system that focuses on internal improvement. This becomes the motivation for information space to improve itself. Specifications for set and multi-dimensional computation provide information space with new processing capabilities for another category of new objects. The growth of information space is separated into individual segments without any logic relations among them and thus becomes a phenomenon attached to the physical or life spaces when looking from the thinking logic of physical and information spaces. Taking another thinking approach, information meaning establishes its own route and law of development in spaces of different nature and under the influences of physical and life laws. The inevitability and regularity of

the disrupted continuous development can be seen. This specific process of information growth is explained in Chapter 5, Section 6.

Chapter 6 is a summary analysis of the information space, consisting of three major parts. One is on the composition of information space and its existence, disappearance and expansion. The other is on the relations between information space and the development of cognitive ability. The third is about the relations between information space and physical and life spaces. As the development of information space is yet to complete a cycle, it is too early to summarize its development patterns. However, three summary points can be observed about the patterns of information space according to its development up to this moment: First, in an independent information space, the quantity of transformation subjects, lasting time, and quality of collaborative platform constitute its expansion scale and development speed, which determines the speed of its growth. We call this the law of lasting collaboration. Second, information space develops in stepped transitions. The transition moves up to a new step after a long, progressive accumulation at every step. We call it the law of accumulative transition. Third, information space is composed according to the principle of non-combination. That is, reflection of corresponding objective existence is done on a one-to-one basis. So is representing the latent or manifest information structure. Copying and processing are also made in a similar manner. This is the characteristic for the composition as well as the growth of information space.

The sixth stage of information development shows the process of information space composition and completion. We must break away from the fixed model that information space is a continuation of the physical world or life world when trying to understand it. It is instead a continuation of information space dominated by laws different from that for matter and life. Laws of the physical space inevitably lead to the emergence of self-existent information and its interaction with organisms. Laws of the life space consequently result in the emergence and evolution of hereditary genes as well as reach the limit of cognitive ability in

organisms. On the other hand, an information space is born which is independent of the physical and life spaces. Its emergence and development also do not follow the physical and life laws.

An important hypothesis is formulated in Section 6.2.5: If information were the key component of the physical space and part of elementary particles or matter particles, it would become the second type of self-existent information in the information space, changing the relations among physical, life and information spaces. Physics neither proves the existence of this state, nor does it prove the non-existence of this possibility. It thus can only be a hypothesis. If this hypothesis is true, information becomes part of the matter in the physical space rather than merely a reflection of the objective existence. Research findings in molecular biology already demonstrate that information is a necessary part of life space. Information is composed of medium, shell and meaning. It takes three states of existence. In the process of information development, the life space is the necessary condition in the crucial stages of information development while the information space relies on the physical and life spaces for its own existence.

The interdependence and interaction of the three spaces create a new outlook of the universe. From the perspective of the universe evolution, the logic of physical space comes from the interaction and unification of energy, conservation of mass, and force. It seeks for unchanging laws in the changeable physical space; The logic of life space, in the unchanging environment, looks for changeable existence and development under the constraint of physical space logic. The logic of information space searches for changes in the operating orbits of physical and life spaces under the constraint of the logic of these two spaces. There are two key aspects in what is described above. One is whether there exists a possibility that the micro-composition and macro-progress of mass is affected via information. The other aspect is regarding who can dominate cognitive ability among the three spaces of physics, life and information. For the first part, physics is still unable to provide an answer today. But it is also invalid to directly negate the possibility. For

the second part, the physical space cannot dominate cognition. The mainstream view believes that the mankind are the subjects of cognition, or at least the subjects that dominate cognitive ability. However, the complexity of realities that cognition is expected to explore and the enhancement of cognitive ability increase at the same speed. The physiological limit of the mankind and rapid growth of autonomous cognitive capability in information space could lead to a problem. If no appropriate solution to the problem is found, it will not be long before non-organisms surpass human beings in cognitive ability. I plan to further discuss and elaborate on this problem in my other book “The Principles of Intelligence” which is to be published in 2017.

This book attempts to clearly define information, the objective existence, and describes the laws of its emergence and development. In addition, this book tries to explore and explain the laws governing the development of the universe through the development of information space and its relations with the physical and life spaces. The current manuscript is a result of research over the past 30 years with numerous negations and affirmations. It is not fully thought over. There is still room for deliberation and improvement. Some conclusions could possibly be wrong. I have it published despite all the limitations in order to obtain readers’ criticisms and improve this theoretical framework.

The preface and table of contents of this book are translated from Chinese into English by Heting Chu, professor at Long Island University, USA. I would like to express my gratitude to her for her help.

I am Grateful and Indebted to Jiuru LIU,Xujun QIN,Yafeng DONG,the editors of this book.

目 录

第 1 章 信息的定义、存在形态和基本属性	1
1.1 信息的定义	2
1.1.1 定义信息必须满足的条件	2
1.1.2 信息的定义	4
1.1.3 什么是, 什么不是	5
1.2 信息的存在形态	6
1.2.1 自在态	6
1.2.2 自有态	7
1.2.3 记录态	9
1.2.4 信息三态之间的关系	11
1.2.5 信息态的转换	13
1.3 信息的一般属性	14
第 2 章 信息的运动	19
2.1 信息形态转换中的运动	20
2.1.1 自在态信息向自有态信息的转换	21
2.1.2 自在态信息向记录态信息的转换	22
2.1.3 自有态信息向自在态信息转换的运动	22
2.1.4 自有态信息向记录态信息的转换	23
2.1.5 记录态信息向自在态信息的转换	24
2.1.6 记录态信息向自有态信息的转换	24
2.2 信息的移动	25

2.2.1	移动的类型	25
2.2.2	信息移动过程的环节与分解	29
2.2.3	信息移动的约束	32
2.2.4	信息移动的演进	35
2.3	信息处理	36
2.3.1	信息处理的定义	36
2.3.2	信息处理的类型	37
2.3.3	转换类信息处理	38
2.3.4	组织类信息处理	41
2.3.5	计算类信息处理	42
2.3.6	信息处理的要素和体系架构	44
2.3.7	信息处理能力的演进	46
2.4	形式信息的积累	49
2.4.1	形式信息积累的定义和类型	49
2.4.2	形式信息积累的要素和架构	50
2.4.3	形式信息积累的目的和路径	52
2.5	载体和外壳的演进	53
2.5.1	载体、外壳、含义的关系及演进模式	54
2.5.2	载体演进和载体与外壳的分离	55
2.5.3	外壳的演进和外壳与含义的分离	57
2.6	本章小结	59
第 3 章	信息结构	63
3.1	信息结构的定义	64
3.2	信息结构的类型和形态	65
3.2.1	自有信息结构和共有信息结构	66
3.2.2	隐性结构和显性结构	67
3.2.3	自在态信息的结构	68
3.2.4	自有态信息的结构	69

3.2.5	记录态信息的结构	70
3.3	信息单元	71
3.3.1	定义	71
3.3.2	信息单元的一般构成	72
3.3.3	信息单元的确立	73
3.3.4	信息结构中几类主要的信息单元	75
3.4	信息单元及信息结构例示	79
3.4.1	二维静态单元的信息结构例示	80
3.4.2	三维静态跨信息态信息结构的示例	84
3.4.3	三维动态跨信息态信息单元的示例	94
3.5	信息结构的标识与函数	100
3.5.1	标识的目的和类型	101
3.5.2	标识的规范和体系	102
3.5.3	函数的类型和特征	103
3.6	信息结构的一般模型	108
3.6.1	信息结构的构成	108
3.6.2	信息结构的一般模型	109
3.6.3	最简信息结构与冗余	112
3.7	信息结构的可用性	114
3.7.1	定义	114
3.7.2	信息结构可用性的一般过程和构成要素	115
3.7.3	隐性结构的可用性和局限性	115
3.7.4	显性结构及其可用性	119
3.7.5	信息结构的显性	121
3.8	显性信息结构的完备性	129
3.8.1	信息结构完备的相对性	129
3.8.2	信息结构完备的测度	132
3.8.3	信息结构间的关系及完备性	137
3.9	关于信息结构的进一步讨论	141

3.9.1	信息结构是连接不同态信息的纽带·····	141
3.9.2	信息结构对加快和引导科学发展具有重要作用·····	143
3.9.3	信息结构的一些重要特征·····	144
第4章	信息的增长·····	147
4.1	信息增长的定义和发展阶段·····	148
4.1.1	信息增长的定义·····	148
4.1.2	信息增长的发展阶段·····	149
4.2	自有态信息的增长·····	152
4.2.1	遗传信息的增长·····	152
4.2.2	认知信息的增长·····	156
4.2.3	非生物自有态信息的增长·····	160
4.3	记录信息的增长·····	161
4.3.1	信息记录的过程·····	161
4.3.2	记录信息增长模型·····	166
4.3.3	记录信息的度量·····	170
4.4	信息结构的生长·····	173
4.4.1	信息结构生长的定义·····	173
4.4.2	显性信息结构生长的一般模式·····	173
4.4.3	显性信息结构生长的基本规范·····	174
4.5	显性信息结构完备的充要条件·····	179
4.6	信息结构显性的承担主体·····	180
4.6.1	主体的属性及动力·····	180
4.6.2	主体的能力·····	181
4.6.3	协同构建与共享·····	183
4.7	显性信息结构完备的先验性一般假定·····	184
4.7.1	一般假定的构成分析·····	184
4.7.2	预期工作量·····	185
4.7.3	复杂性·····	187

4.7.4	实施路径	190
4.7.5	信息结构显性完备的动力来源假定	191
4.8	信息结构显性的方法和工具	192
4.8.1	方法论的一般讨论	192
4.8.2	工具的形成	194
4.8.3	开放、开源和交互平台	201
4.8.4	工具的演进	204
4.8.5	长期进化	207
4.9	信息增长的规律和轨迹	208
4.9.1	独特的增长过程和规律	208
4.9.2	跨越转折点	210
4.9.3	创建信息处理新逻辑	211
第 5 章	信息的逻辑	215
5.1	信息逻辑的架构	216
5.1.1	信息按不同规律运动的起源	216
5.1.2	信息逻辑的组成要素	219
5.1.3	信息逻辑分析的标识方式	220
5.2	感知、转换和映射	222
5.2.1	信息含义在信息转换和移动中的一般规律	222
5.2.2	生物体的感知和转换	224
5.2.3	信息系统的感知或转换	227
5.3	连接场及其表述	229
5.3.1	连接的对象	230
5.3.2	连接的形式	235
5.3.3	连接的实现	237
5.4	计算规划和架构	240
5.4.1	计算规划与架构的一般性问题	240
5.4.2	微处理、短逻辑	243

5.4.3	泛包容	244
5.4.4	内计算	246
5.4.5	多中心计算	248
5.4.6	结构的持久扩展	249
5.5	构件集合的计算和多维计算	251
5.5.1	构件集合的比较	251
5.5.2	基于构件集合的组合和分解	253
5.5.3	信息结构的维度和多维计算	254
5.6	非连续的连续进化	255
5.6.1	非连续的连续进化规律	256
5.6.2	感知力	258
5.6.3	表达力	259
5.6.4	解释力	261
5.6.5	重现力	262
5.6.6	中断与连续	263
5.6.7	进化的时空逻辑	265
5.7	信息逻辑的进一步讨论	267
5.7.1	连接和计算逻辑的一般问题	267
5.7.2	显性信息结构自我进化的逻辑问题	268
5.7.3	信息逻辑的基本特征	269
第 6 章	信息空间	271
6.1	信息空间	272
6.1.1	信息空间的构成	272
6.1.2	信息发展阶段及信息空间构成的变化	274
6.1.3	主体的演变	276
6.1.4	信息空间的存在、消亡与扩展	278
6.1.5	信息空间进化的几个规律	280
6.2	信息空间与认知能力发展的关系	282

6.2.1	认知能力与信息空间发展的一般关系	283
6.2.2	信息空间发展初期的认知能力	285
6.2.3	信息空间发展中期的认知能力	286
6.2.4	信息空间发展后期的认知能力	289
6.2.5	一个假设：如果信息是物理世界的构成要素	290
6.3	信息空间与物理空间和生命空间的关系	292
6.3.1	信息空间与物理空间	293
6.3.2	信息空间与生命空间	293
6.3.3	三个空间互动下的宇宙演进	294

Contents

Chapter 1	Definition, Existence States and Basic Attributes of Information	1
1.1	Definition of Information	2
1.1.1	Requirements for Defining Information	2
1.1.2	Information: A Definition	4
1.1.3	What is Information, What is not Information	5
1.2	Existence States of Information	6
1.2.1	The Self-existent State	6
1.2.2	The Self-containing State	7
1.2.3	The Recorded State	9
1.2.4	Relations Among the Three States of Information	11
1.2.5	Transformation of Information States	13
1.3	Basic Attributes of Information	14
Chapter 2	Motion of Information	19
2.1	Motion in Transformation of Information States	20
2.1.1	Motion of Information Transformation from the Self-existent State to the Self-containing	21
2.1.2	Motion of Information Transformation from the Self-existent State to the Recorded	22

2.1.3	Motion of Information Transformation from the Self-containing State to the Self-existent	22
2.1.4	Motion of Information Transformation from the Self-containing State to the Recorded	23
2.1.5	Motion of Information Transformation from the Recorded State to the Self-existent	24
2.1.6	Motion of Information Transformation from the Recorded State to the Self-containing	24
2.2	Move of Information	25
2.2.1	Types of Information Move	25
2.2.2	Analysis of Steps in the Process of Information Move	29
2.2.3	Constraints on Information Move	32
2.2.4	Development of Information Move	35
2.3	Information Processing	36
2.3.1	Definition of Information Processing	36
2.3.2	Types of Information Processing	37
2.3.3	Transformational Information Processing	38
2.3.4	Organizational Information Processing	41
2.3.5	Computational Information Processing	42
2.3.6	Essentials and Framework of Information Processing	44
2.3.7	Development of Information Processing Capacity	46
2.4	Accumulation of Form Information	49
2.4.1	Definition and Types of Form Information Accumulation	49
2.4.2	Essentials and Framework of Form Information Accumulation	50
2.4.3	Objective and Approach of Form Information Accumulation	52
2.5	Development of Information Media and Shells	53

2.5.1	Relations Among and Major Developments of Information Medium, Shell and Meaning	54
2.5.2	Information Medium Development and the Detachment of Medium and Shell	55
2.5.3	Information Shell Development and the Detachment of Shell and Meaning	57
2.6	Chapter Summary	59
Chapter 3	Information Structures	63
3.1	Definition of Information Structure	64
3.2	Types and States of Information Structure	65
3.2.1	Self-containing Information Structure and Shared Information Structure	66
3.2.2	Latent Information Structure and Manifest Information Structure	67
3.2.3	Self-existent Information Structure	68
3.2.4	Self-containing Information Structure	69
3.2.5	Recorded Information Structure	70
3.3	Information Units	71
3.3.1	Definition of Information Unit	71
3.3.2	Common Components of Information Unit	72
3.3.3	Establishment of Information Unit	73
3.3.4	Major Types of Information Unit in Information Structure	75
3.4	Examples of Information Units and Information Structures	79
3.4.1	Example of Two-dimensional Static Information Unit Structures	80
3.4.2	Example of Three-dimensional Static Cross-state Information Structures	84
3.4.3	Example of Three-dimensional Dynamic Cross-state Information Units	94

3.5	Representation and Mathematical Functions of Information Structure	100
3.5.1	Representation Purposes and Types	101
3.5.2	Specifications and Frameworks for Representation	102
3.5.3	Function Types and Features	103
3.6	Composition and General Model of Information Structure	108
3.6.1	Composition of Information Structure	108
3.6.2	General Model of Information Structure	109
3.6.3	Simplest and Redundant Information Structures	112
3.7	Applicability of Information Structure	114
3.7.1	Definitions for Information Applicability and Information Structure Applicability	114
3.7.2	General Process and Major Components of Information Structure Applicability	115
3.7.3	Applicability and Limitations of Latent Information Structure	115
3.7.4	Applicability of Manifest Information Structure	119
3.7.5	Manifestation of Information Structure	121
3.8	Completeness of Manifest Information Structure	129
3.8.1	Relativity of Information Structure Completeness	129
3.8.2	Measures for Information Structure Completeness	132
3.8.3	Relations Among Information Structures and Completeness	137
3.9	Further Discussion on Information Structure	141
3.9.1	Information Structure as Link Connecting Information of Different States	141
3.9.2	Significant Impact of Information Structure on Facilitating and Guiding Scientific Development	143
3.9.3	Six Important Characteristics of Information Structure	144

Chapter 4	Information Growth	147
4.1	Definition and Development Stages of Information Growth	148
4.1.1	Definition of Information Growth	148
4.1.2	Development Stages of Information Growth	149
4.2	Self-containing Information Growth	152
4.2.1	Hereditary Information Growth	152
4.2.2	Cognitive Information Growth	156
4.2.3	Non-biological Self-containing Information Growth	160
4.3	Recorded Information Growth	161
4.3.1	Process of Information Recording	161
4.3.2	Recorded Information Growth Model	166
4.3.3	Measuring Recorded Information	170
4.4	Information Structure Growth	173
4.4.1	Definition of Information Structure Growth	173
4.4.2	General Model of Manifest Information Structure Growth	173
4.4.3	Basic Principles and Specifications for Manifest Information Structure Growth	174
4.5	Prerequisites and Requirements for the Completeness of Manifest Information Structure	179
4.6	Subjects Responsible for Manifesting Information Structure	180
4.6.1	Attributes and Motivations of Subjects	180
4.6.2	Abilities of Subjects	181
4.6.3	Collaborative Construction and Sharing	183
4.7	A Priori, General Assumptions about Manifest Information Structure Completeness	184
4.7.1	Composition Analysis of General Assumptions	184
4.7.2	Expected Workload and Resources	185

4.7.3	Complexity	187
4.7.4	Implementation Approaches	190
4.7.5	Assumptions about Motivation Sources for the Completeness of Information Structure Manifestation	191
4.8	Methods and Tools for Information Structure Manifestation	192
4.8.1	General Discussion on Methodology	192
4.8.2	Formation of Tools	194
4.8.3	Crowdsourcing, Open Source and Interactive Platforms	201
4.8.4	Development and Enhancement of Tools	204
4.8.5	Long-term Efforts in Tool Improvement	207
4.9	Patterns and Paths of Information Growth	208
4.9.1	Unique Growing Process and Patterns	208
4.9.2	Crossing the Turning Point of Information Growth	210
4.9.3	Establishing New Computing Logic for Information Processing	211
Chapter 5	Information Logic	215
5.1	Framework of Information Logic	216
5.1.1	Origin of Information with Different Motion Patterns	216
5.1.2	Major Components of Information Logic	219
5.1.3	Representation Methods of Information Logic Analysis	220
5.2	Perception, Transformation and Mapping	222
5.2.1	General Patterns of Meaning in Information Transformation and Move	222
5.2.2	Perception and Transformation of Organisms	224
5.2.3	Perception and Transformation of Information Systems	227
5.3	Connection Field and Representation	229
5.3.1	Connection Objects	230
5.3.2	Connection Forms	235

5.3.3	Connection Implementations·····	237
5.4	Computation Planning and Framework ·····	240
5.4.1	General Problems in Computation Planning and Framework ·····	240
5.4.2	Micro-processing and Short Logic ·····	243
5.4.3	Universal Inclusion ·····	244
5.4.4	Internal Computation ·····	246
5.4.5	Multi-center Computation ·····	248
5.4.6	Sustained Expansion of Structure ·····	249
5.5	Component Set Computation and Multi-dimensional Computation ···	251
5.5.1	Comparison of Component Sets ·····	251
5.5.2	Component-set-based Combination and Decomposition·····	253
5.5.3	Dimensions and Multi-dimensional Computation of Information Structure·····	254
5.6	Disrupted Continuous Development·····	255
5.6.1	Law of Disrupted Continuous Development·····	256
5.6.2	Perception·····	258
5.6.3	Expression ·····	259
5.6.4	Explanation·····	261
5.6.5	Representation ·····	262
5.6.6	Disruption and Continuation ·····	263
5.6.7	Temporal and Spatial Logic for Development ·····	265
5.7	Further Discussion on Information Logic·····	267
5.7.1	General Problems in Connection and Computation Logic ···	267
5.7.2	Logical Problem in Self-development of Manifest Information Structure·····	268
5.7.3	Basic Characteristics of Information Logic ·····	269

Chapter 6	Information Space	271
6.1	Information Space Overview	272
6.1.1	Composition of Information Space	272
6.1.2	Information Development Stages and Subsequent Changes in Information Space Composition	274
6.1.3	Changes of Subjects in Information Space	276
6.1.4	Existence, Disappearance and Expansion of Information Space	278
6.1.5	Some Laws for Information Space Development	280
6.2	Relations Between Information Space and Cognitive Ability Development	282
6.2.1	General Relations Between Cognitive Ability and Information Space Development	283
6.2.2	Cognitive Ability in the Early Stage of Information Space Development	285
6.2.3	Cognitive Ability in the Middle Stage of Information Space Development	286
6.2.4	Cognitive Ability in the Late Stage of Information Space Development	289
6.2.5	One Hypothesis: If Information were the Key Component in the Physical World	290
6.3	Relations Between Information Space and Physical and Life Spaces	292
6.3.1	Information Space and Physical Space	293
6.3.2	Information Space and Life Space	293
6.3.3	Development of the Universe Via the Interaction Among the Three Spaces	294

第 1 章

信息的定义、存在形态和基本属性

我们感觉这个世界的信息似乎越来越多，但意义却越来越少。
意义强行回归是不可避免的。

——[法] Jean-Pierre Dupuy¹

印在纸上的四分和八分音符不是音乐。音乐既不是空气中的一串压力波，也不是刻在胶木唱片上的沟纹或烧在光盘上的凹洞，更不是搅动听众的大脑神经元交响曲。同样，DNA 的碱基对不是基因，它们为基因编码，基因是由比特构成的²。

这座图书馆（指虚拟的收录了所有图书的图书馆）永远存在，它就是宇宙。至于我们，还没有写下一切，也没有变成幻影。我们走过长廊，在书架上搜索，并将其重新排列，在众多的杂音和不一致中寻找几行有意义的信息；阅读过去和未来的历史，收藏自己和别人的想法，一次次看向镜中，或许我们每一瞥都能看到信息创造的生物。

——[美] James Gleick³

信息就是意义，信息将构造一个新的宇宙。这是本书的起点和终点。

1 The Mechanization of the Mind: On The Origins of Cognitive Science, [法] Jean-Pierre Dupuy, M.B.DeBevoise 译 Princeton, N.J.: Princeton University Press, 2000.

2 见 James Gleick, The Information, a History, a Theory, a Flood, Fourth Estate, London, 2011, p.309.

3 见 James Gleick, The Information, a History, a Theory, a Flood, Fourth Estate, London, 2011, p.426.

1.1 信息的定义

控制论的创始者，被称为信息时代之父⁴的诺伯特·维纳认为，“信息就是信息，不是物质，也不是能量”。⁵与物质和能量相提并论的信息究竟应该是什么？

1.1.1 定义信息必须满足的条件

信息一词已经渗透到各个方面。[美]格莱克著，2011年出版的《信息》⁶一书，从非洲会说话的鼓声，到文字的产生、字典式百科全书、计算机、信息网络、通信和控制的信息理论、热力学第二定律、遗传基因、社会文化的继承到今天爆发式增长的信息洪流，洋洋洒洒，描述了数万年信息演进历史，试图从各方面剖析信息、认识信息，但还是遗漏了一些重要的信息类型，例如，大脑和神经系统处理的认知信息、量子信息，也没有给出信息的明确定义。应该肯定地说，对于今天几乎所有人、所有领域都在使用的、作为信息社会基石的信息一词还没有具有共识的定义。

几十年来，基于不同的学术背景，对信息提出了数百种的定义。历来，对于信息的定义最宽泛的是所谓本体论的定义，《中国大百科全书》（1993年）对信息词条的解释为：信息泛指一切事物运动的状态和运动的方式，包括事物内部结构的状态和方式以及外部联系的状态和方式；从认识论意义上说，信息是关于事物运动状态和运动方式的反映。这个定义覆盖的范围很广，但

4 Dark hero of the information age, in search of Norbert Wiener, the father of Cybernetics, Flo Conway and Jim Siegelman, 2005 年，这是一本维纳的传记，第一句话就是：He is the father of the information age.

5 Norbert Wiener: Cybernetics: or Control and Communication in the Animal and the Machine, New York, John and Sons, 1948.

6 The Information, a History, a Theory, a Flood, [美]James Gleick, Fourth Estate, London, 2011.

解释性、规范性都不够严密。例如，“反映”是过程还是结果？还是既是过程又是结果？尤其是，这一定义没有将信息与物质、能源和生命区分出来，运动的状态和方式可以是物质的、生命的，也可以是信息的，这一定义实际上没有在内涵和外延两个关键维度界定信息。

影响力最广泛的定义是以通信或控制过程消除的不确定性，格莱克所著《信息》一书，用了两章介绍通信系统中处理的信息，核心就是申农通信系统中的数学理论，后人衍生为“信息论”。美国数学家申农（C.E.Shannon）1948年发表在《贝尔系统技术杂志》的一篇论文中，在讨论通信过程中信源信息量和信道容量的计算方法时，将信息熵定义为“随机不定性之差”，是用于消除随机不定性的东西，在随后的一篇论文《通信的数学理论》⁷中有效地建立了电信系统中信息传输的优化模型，但他自己明确声称，从来没有研究过信息的意义。显然，没有包括信息的含义，没有包含其他信息外形的运动方式，申农所指信息不是信息的恰当定义。

控制论的主要创始人维纳（Nobert Wiener）则认为“信息就是我们适应外部世界，并把这种适应反作用于外部世界的过程中，同外部世界进行交换的内容的名称”⁴。显然，维诺的定义与申农相反，重在意义。如果把这称为定义，可能维诺也不同意，因为同其强调的“信息就是信息，不是物质，也不是能量”的设想相去甚远。即使是肯定这样的定义，也只包含了信息运动的部分类型和各种应该包含在信息中的经典类型，如百科全书、基因等。

此外还有不少信息的定义，不是只针对特定的信息类型，就是只定义特定的信息运动过程，有的则语焉不详、所指不确定，例如定义为某种有用性、负熵等。这样的定义当然不能取得共识。

定义信息，至少满足三个要求，一是解释性，对林林总总的称之为信息的实体及其发生和发展具有解释能力，能统一归到这一定义下；二是规范性，满足信息定义的应该是具有共同特征的客观实在；三是等价性，定义的信息要与物质、能源和生命等在概念范畴和实际影响力上具有相似性。下边的文

7 The Methematical Theory of Communication, Claude Elwood Shannon and Warren Weaver, Urbana, University of Illinois Press, 1949.

本框通过一个几乎每个人可以体会的实例，说明了定义信息解释性规范性要求的目的和意义。

一片叶子的信息旅行，解释性和规范性要求例子。

我家十几年来，养了几盆四季报春花。稍有余闲，透过叶片看生长是否正常，这是该花在光线的作用下，不管你观察与否，无时无刻不在散发的信息。每年5月初，我会摘下几片叶子，或切几段带叶的茎重新种植，没几天，叶子就会长出根茎，萌出新叶，不断长大，到12月开花，5月再重栽，周而复始。这是遗传信息在发挥作用。如果我把看到的过程记录下来，或用手机拍个照，录个像，则又产生了两个重要的信息类型，一是在大脑构思时形成，并在一定阶段保存在大脑中的信息；二是记录下来的书面信息，不管是书写在纸上，还是记录在电脑或手机中。如果在栽种时，同时对该花的叶片进行基因组排序，把这种花的遗传基因及遗传机制用模拟和数字的方法科学、忠实地记录下来，那么这个过程将存在三种基本等同的信息，植物本身的生长过程信息、人大脑中的信息和记录下来的信息。如果使用计算机进行处理，则还有以电路开关（0或1）的方式存在，但在含义上与前述三种信息等价的机内信息。

1.1.2 信息的定义

定义 1.1：信息。满足上述要求的定义是：信息是由载体与/或外壳承载并表达的一切客观存在的含义。这里含义是意义的同义词，对应的英文词是 Meaning。对这一定义要说明五点：第一，信息是含义，是内容，不是物质、能量、生命等任何承载；它的载体，也不是文字、声音、图像、核苷酸、神经元等任何表示它的形式性外壳，但载体和形式性外壳又是信息存在的必要条件。第二，这里的客观存在包含物质、能量、生命和信息本身，因为如此

定义的信息也是一种区别于物质、能量和生命，又不为前述三者包含的另一类客观存在。第三，这里所指客观存在，既包括自然界，也包括人类社会；既包括实体，也包括过程；既包括个体，也包括群体；既包括静态存在，也包括各种简单或复杂形态的动态事件；既包括以信息形态表达出来的，还包括将要以信息形态表达出来的，这里的“将要”没有时间长度的约束。第四，这里所指客观存在包括生物或非生物智能体创造物或生成物，不管这种创造物或生成物是以物质的方式存在还是以信息的方式存在。所谓生成物是指由智能体产生，但在有些定义中不构成创造的东西，如意识、记忆。第五，这里所指表示的客观存在含义，不仅是指其属性，还包括其功能；不仅包含功能是什么的含义，还包括功能是如何实现的含义。

每一个人都有很多自身经历的体会，说明自己心中所关注的信息是含义，而不是载体或外壳。例如，不管是听到信息、资讯、情报、Information，还是 информация，脑子里存储和理解的不是声波本身，而是相应声波后面不同外壳下共同的“信息”这一含义。同样，我们看到上述这些单词，接受的不是眼睛感受的光及其组成的不同字形，而是共同的含义。

1.1.3 什么是，什么不是

根据这样的定义，考虑到已经存在的大量信息定义或在实践中使用着的信息概念，需要界定什么属于，什么不属于本书定义的信息。

第一，一切客观存在本身不是信息，只有这样的存在以一定的载体与/或外壳展示其含义的才是信息。一切客观存在包括信息，但关于信息存在的信息不是指信息所指代的含义，而是信息的载体与/或外壳存在状态的含义。

第二，物理或生命的过程与/或度量本身不是信息，而是存在；但过程的反映和描述，度量过程及结果的描述、记录、方法及结果符合信息的定义。克劳修斯的热力学第二定律阐述的熵增加过程，申农通信的数学理论中通信

过程及消除不确定性的过程均不包含于这一定义的信息范畴内⁸。

第三，量子信息一词若不是用来指物质构成要素，而是指量子的状态测量、量子状态用作通信或信息处理新型载体的量子信息不包含在本文的定义中。

第四，如此定义的信息容纳了除上述三类外几乎所有习惯使用的信息类型。存在于广泛物理空间的以光、声、电磁波辐射、散布的信息；存在于生命体的遗传信息，生命体神经系统感知、传输、存储、处理的信息；人类社会及其他生物体的生存发展状态信息，人及其他智能生物体的语言或准语言；人类或其他信息处理系统收集并记录下来的信息；人类或其他信息处理系统（从图书馆到互联网）不同层次加工整理存储的信息，科学、技术、文化等知识信息，娱乐、衣食住行等生活信息。

1.2 信息的存在形态

含义、载体和形式性外壳结合在一起，成为信息的实际存在形态，信息的存在形态丰富多样。为便于研究信息发生和发展规律、更好地利用信息，根据其自身特点，信息跨入分成自在、自有和记录三类存在形态。

1.2.1 自在态

作为客观存在的事物，除了事物本身，还以信息的方式，向外部世界显示其存在和特征。物质以反射光、声波、辐射、引力及其他作用力显示其存在和特性；植物以其生长和与周边的互动显示其存在和特性；动物则以丰富

8 热力学第二定律描述的一个封闭空间无序状态的度量值显然不属于本书定于的信息范畴，因为这是一个物理过程的结果，反映或描述这个物理过程一切内容，包括存在于不同载体的，也包括度量方法、过程及结果，都是信息，而不仅是作为结果的度量值。香农关于信息的有关讨论源自通信的数学理论，是通信过程不确定性的度量。香农自己认为，在他的研究领域，“信息的意义无关紧要”，“在本文中，信息一词虽然与其惯常的词义有关，但不应该与之混淆”。而本书所定义信息的本质就是含义，所以和香农关于信息一词的使用南辕北辙。

得多的活动信息与周边交互，相互影响；人类社会个体和群体的每一活动、事件均自在地发出信息。

定义 1.2: 自在态信息。自在态信息是指一切客观存在物，包括非生物体个体、非生物体群体、生物体个体和生物体群体在物理空间遵循物理规律，自行散发出的信息。自在态信息与物质世界同生同在。时空无限，自在态信息无限；物质世界的物理力及其相互作用无处不在，所以自在态信息弥漫性散布于物理空间中。自在态信息与生命世界同生同在。生命世界多姿多彩，自在态信息丰富多样。

一切智能体创造物也是客观存在物，关于它的存在和运动也属于自在态信息。由于关于该创造物的自身内含的信息一般已经以记录方式存在，属于记录信息。如关于一架飞机，那么相关的信息，如设计文档、制造工艺说明、维护手册等均为记录信息，而不是自在态信息；至于非物质创造物，如文学、艺术、科技著作等无疑是记录信息。

没有信息，人们观察不到事物的本来面目和运动状态、规律。自在态信息不能自然地转换为自有信息或记录信息，没被转换的信息，意味着在含义层面没有被使用，只是一种客观存在，没有进入信息增长和演进过程。不同的自在态信息转换为自有或记录信息，需要不同的条件和过程，自在态信息转换为自有或记录态信息是信息的运动，在下一章讨论。

1.2.2 自有态

定义 1.3: 自有态信息。自有态信息是指各种物体内含的信息。自有态信息有两大类：一是非生物体和生物体自身不能进行含义运动变化为目的进行处理的内含信息，所有非生物体的内含信息均属于这一类；二是生物体、机器智能、信息系统自身可用的信息，即自有且自用的信息。由于第一类不能直接进入信息增长和演进的链环，本书以下部分在没有特别注明的情况下，自有态信息特指第二类。

第二类自有态信息根据其使用的方式，还可以划分为两个大类，六个小类。一大类是使用过程不与外部存在任何信息交互的自有信息；另一大类是

在使用过程中与外部存在不同形式的信息交互的自有信息。

六个小类分别为：第一类，遗传基因是一类特殊的自有信息，保证着生命的延续和进化，既具有独特的增长和演进路径，更在信息的增长和演进历史中扮演重要角色，基因工程可以恰当地解释后一种作用。科学地揭示基因内含信息，通过大脑的作用并记录下来积累，与人类已经形成的其他技术和科学一起，通过基因工程改变基因，推动了生命的演进，同时也成为信息增长和演进的一个具体过程。

第二类，仅在生物体内发生作用，但没有继承和积累的自有态信息，就是生物体自身维持生机的功能性信息，产生本能反应的信息。这类信息是生物体除智能活动和遗传基因之外的自有态信息，本质上是自在态信息，是生命科学研究对象，本书不再作深入讨论。

第三类，语言是又一类特殊的自有态信息。当我们把语言（一切作为生物群体可理解的交流工具的声音）看作物理空间的声波，它是自在态信息，当我们把这样的声音录下来，则形成了记录信息，但更应该把它看作该生物群体之间交流的信息，即群体性自有态信息。人与人之间口头交流的语言是其中迄今为止最完备的语言性自有态信息。任何智能生物群体间没有记录下来的交流信息，都是群体性自有态信息。

第四类，生物的认知性自有态信息，是关于智能生物智能行为和智能发展的自有态信息。智能生物经由感知、学习、探索，高等智能生物（例如人）还可经由系统性研究，其神经系统和大脑使用、存储、处理并与外界交互的自有态信息。对这类信息，重要的是把握同样含义，却分属自在、自有、记录三大类型。如，一个人用天文望远镜观察北斗七星，记录云彩和风对观察的影响。那么，观察视野中北斗七星、其他星空、云彩和风在光的作用下散发出自在态信息；观察者经由天文望远镜感知自在态信息，经由神经系统依据既有的相关框架，在大脑中形成图像和描述的自有态信息；在根据这样的自有态信息，写下观察记录，称为记录态信息。生物的智能性自有态信息是信息增长和演进的关键环节。

第五类，信息系统的自有态信息。这里的信息系统特指无智能的人造的由相关信息技术设备构成的用于感知、传输、处理、存储信息的系统。因为

该系统是无智能的，所以一切信息处理均根据人设定的指令操作。这样的信息系统的自有态信息就是存在其中，系统可处理的信息。其中可处理的约束十分重要，不是可处理，那就成为图书馆、藏书楼了，成为记录信息的载体。这类信息在性质上与第二类信息相近，变化源自外在的人的指令，而不是系统本身。

第六类，非生物智能体（机器智能）的自有态信息。机器智能也是一种信息系统，只是具有自主的具有智能特征的信息系统，如机器人、具有一定复杂性的自动化系统、智能生产或指挥系统等。这类自有态信息是指系统中不同于第五类所描述的信息行为特征，可以具有如第四类规定的信息行为特征，即可以自我获取信息，自我发起处理，形成新的存在与系统中的信息。

1.2.3 记录态

定义 1.4: 记录态信息。记录态信息是指对自在态信息和自用态信息通过一定的工具、用一定的载体记录下来，并能够为特定智能体群理解和使用的信息。记录态信息是三种形态中最年轻的，经历了近万年持续发展，从结绳记事到维基百科，从泥板印刷到激光照排，从极少数人具备的能力到数十亿人和更多数量的设备参与信息的记录，记录态信息快速增长，信息爆炸这个词指的就是记录态信息。记录态信息表观形式繁多，录音、录像、书籍、磁介质，文字、数据、音视频等都是记录态信息。记录态信息是跨时空延续的必要基础，是信息形态增长和演进的最高形式，是智能进化的结晶。按照记录态信息成因，可分为自在态信息的直接记录和经由人类大脑或非生物智能体加工后记录的信息。而后者又可以细分为业务性、学术性、生活性三类，之所以按这样的标准对全部记录态信息分类，是为了有利于对信息增长和演进规律的剖析。

第一，自然世界信息的直接记录。从太空望远镜到海底声纳，从各种光谱感知到震动勘探，从手机到专业摄影录像机，从台风路径到雾霾分布，从熊猫出没到太阳风暴，人类在越来越多地记录自在态信息。随着感知技术的进步和应用的扩展，自在态信息的直接记录信息数量已经并将继续急剧增长，

论信息

成为记录态信息的主要组成部分。自在态信息的直接记录是信息发展过程的中间环节，直接记录的目的是用于人类社会发展和生存，对应于国民经济和社会各个领域发展需求、研究和创新需求、个体和社会的生存与发展需求、余暇和娱乐等休闲活动需求。

第二，业务性记录态信息。这里，业务一词是泛指，政治、经济和社会发展、公共管理、国防军事、文化体育等一切事务都包含其中。将业务信息尽可能完整地记录下来，是管理的基本要求，是进步和发展的基础资源。企业的设计生产、经营管理，政府的公共服务、社会管理，科研教育、文化卫生的管理和事务，国防、安全事务等在业务工作中记录下来的信息均属于此。业务记录信息在其不同子类中存在重复、交叉。如企业统计信息成为政府统计或其他宏观管理业务的组成部分，金融业务数据成为税务工商部门的业务信息。业务记录信息与直接记录信息之间存在交集。如自动控制设备、监控音视频等直接记录信息不经智能体的修改成为业务记录信息。业务信息与学术信息存在交集。各种业务信息，特别是具有空间或时间系统性的业务信息，都可以作为研究的基础，有些并被直接转入学术信息范畴，但本质上还是业务信息。

第三，学术性记录态信息。人类自有记录信息以来积累的各个知识门类的信息总称为学术性记录态信息。把自己的研究成果用一定的方式公之于众，是学者的共同心态，是社会进步的基本动力。学术信息是人类社会智能进化的结晶（到今天为止是如此，但当非生物智能体介入智能演进，并形成学术成果时，应称作智能进化的结晶），是人类认识世界、改造世界成果的积淀，是文明演进的基础。数千年来，人类持之以恒的努力探索，形成了今天包罗万象、体系化的知识体系，这是学术信息的浓缩版、精化。但是，人类在把物质世界、人类社会和人的研究取得不断深化的结果时，学术信息的研究却没有取得应有的成果，知识越多，学习和研究越困难。要像对其他客观存在研究的精神和投入来研究学术信息这一类特别复杂的客观存在，使之成为加速研究创新的推进剂，而不是相反。本书第3章和第6章将重点讨论这一问题。

第四，生活性记录态信息。人们关注自己，这是天性。工作属于集体、

生活属于自己，当条件具备时，人们必然记录自己的、家人的、朋友的、社群的生活信息，这是生活性记录信息发生发展的动力。在文字发明的早期，记录信息大都与生活信息有关，祭祀、天气、收成，随着国家的形成和发展，加上记录保存信息的成本和记录本身的能力要求，记录信息的主体转向宫廷和富人，在留存的记录中，生活信息的比重减少了。到了 21 世纪，教育的普及，人均收入水平的提高，互联网渗透到世界各地，音视频记录设备价格的降低，四大因素融会在一起，生活信息的记录急剧增长，数量上已经超越学术性信息，超越业务性信息也是必然的，只是时间问题。在生活性记录信息中，有的需要保留，有的即记即用，没有后续使用必然性，这些特征将在讨论信息的运动和增长是进一步分析。生活性记录信息可能同时也是直接记录信息，有的部分也可能是学术性、业务性记录信息的构成部分。

1.2.4 信息三态之间的关系

从外延看，自在态、自有态、记录态这三种类型的信息构成了信息的全集，但这三态之间存在不可分割的联系，联系的基础是含义，联系的纽带是转换，联系的中枢是显性信息结构。如果说感知的才是存在的，那么没有内涵，就没有外在；如果说存在才被感知，那么外在对内含具有本质性的反作用，具体的自有信息与自在和记录信息存在密不可分的复杂关系。信息三态如图 1.1 可以划分为 10 个区域，定义这 10 个区域，就是对三态间关系的一种解释。

图 1.1 的三个圆分别表示自在、自有和记录信息，这三个圆确定了前述定义的信息外延。虚线构成的圈为所有符合自在信息定义的集合 Z ，用虚线表示无限之意；自有信息箭头所指之圈为所有符合自有信息定义的集合 Y ；记录信息箭头所指之圈为所有符合记录信息定义的集合 L 。

集合 A 为集合 Z 减去与集合 Y 和集合 L 的交叉部分，表示自在信息没有为自有信息感知记忆，也没有为记录信息留存的部分。

集合 B 为集合 Y 减去与集合 Z 和集合 L 的交叉部分，表示按定义属于自有信息而不属于自在或记录信息，即纯粹的自有信息，既没有对应的可感知客观实在，也没有记录下来，成为记录态信息，例如人或其他智能体自己的

推论、想象、假定、猜测等，这部分信息几乎每个智能体或智能体群体间的口头交流中都有存在。没有文字的智能物种的自有信息中，去除遗传信息，均属于这个集合。

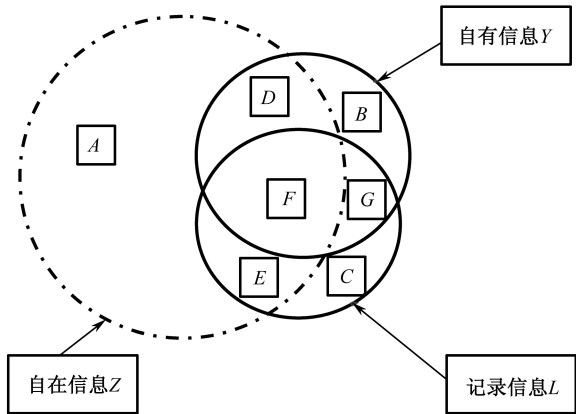


图 1.1 信息三态及其关系

集合 C 为集合 L 减去与集合 Z 和集合 Y 的交叉部分，表示按定义属于记录信息而不属于自在或自有信息，即纯粹的记录信息，既不是自在信息的真实反映，也不是自有信息的真实记录，也就是已经记录但并不源自于自有信息，也不是正确的客观实在的反映。

集合 D 是集合 Z 和集合 Y 的交集，表示其中的每一条信息既是自在信息又是自有信息，但不是记录信息，一信息之所以归入该集合，显然与承载该信息的物理载体无关，生物智能体神经网络中存在，而没有记录的部分应属于这个集合。

集合 E 是集合 L 与集合 Z 的交集，表示其中的每一条信息既可归入自在信息又可归入记录信息，而又不是自有信息，一信息之所以归入该集合同样与承载该信息的物理载体无关。归入这个集合信息主要表示已经记录下来，而生物智能体还没有理解或知悉的信息，大体上属于各类信息获取设施记录下来，而还没有处理或没有能力转化为生物智能体可以理解使用的信息。

集合 F 是 Z 、 Y 、 L 三个集合的交集，其中的每一个元素按定义可归入所

有三大类信息。一条信息同属自在、自有、记录，显然与承载该信息的物理载体无关，证实了信息是客观存在的含义这个定义。这一集合的信息横跨三大类型，体现了三个类型信息的特征，对研究信息结构、增长、度量等方面具有特别的意义。

集合 G 中的每一个元素按定义属于集合 Y 和集合 L ，但不属于集合 Z 。归入这个集合的信息意味着智能主体思考、反映并记录下来，但这些信息不符合真实的客观存在或不是客观存在的反映，很多文学作品属于这个集合。

分析这 10 个信息区域，进一步从外延上定义了信息，基本界定了信息三态之间的边界，但三态之间还有一些其他重要关系需要说明。一是从宇宙演进的历史看，在这三态中，自在信息发生最早，是自有和记录信息的基础，换言之，从历史看，没有自在信息，就没有自有信息和记录信息。关于这一问题，在下一节还要进一步分析。二是从数量看，自在信息与时空同在，既有限的存在，又有无限的冗余。自有信息是生物体和非生物智能体容量和数量的函数，但不管数量与容量如何增长，总是有限集合，但在数量上大于记录信息，即使记录信息在传感技术和人类社会高度重视的历史条件下不断增加，总量上依然比自有信息少很多。三是三者之间联系的基础是含义。上面对集合 F 的分析已经说明了这一点，如果信息不是含义本身，而是与特定的载体相连，那么集合 F 就不存在因为集合 Z 、 Y 、 L 的载体和外壳是个不相同的，是不同的载体和外壳承载着相同的含义，使得集合 F 变为研究信息发生发展规律最重要的部分。四是三者之间联系的纽带是转换，存在客观的可观察、可度量、可重复的转换模式。如果三者之间不存在转换的通道，图 1.1 就失去了意义，集合 Z 、 Y 、 L 就是三类独立的、不相交的客体。

1.2.5 信息态的转换

信息态转换是指含义不变，载体与/或外壳发生了变化。信息三态间的转换是信息运动的重要组成部分，是信息发生发展规律的一个重要内容，存在如下八种转换形式。

第一，自在态信息转换为自有态信息。例如人在走路时对周边观察经由视觉、听觉或触觉进入大脑，并在大脑中构成周边环境的图像，为下一步怎

论信息

么走这一决策服务。

第二，自在态信息转换为记录态信息。例如城市道路监控、小区安全防护监控、自动化制造设备或生产线的作业状态的监控反馈，都属于这一转换类型。

第三，自有态信息转换为记录态信息。例如自动化生产线根据程序设定的要求，将作业进展打印出来，递交给管理人员。再如科研论文的写作，不管是博士论文还是学术会议宣读的论文，都是自有态信息转换为记录态信息的典型案例。

第四，自有态信息转换为自在态信息。例如，动植物生长过程中遗传信息决定动植物形状和生长主要特征的过程。遗传基因是生物体一类最重要的自有信息，通过一系列复杂而有规律的转录、复制、依据外在环境对生长的控制，成功地生长出一只具体的动物或一枝具体的植物，这样的客观存在，又是自在态信息的来源。

第五，记录态信息转换为自有态信息。阅读是记录态信息转换为自有态信息的典型场景。

第六，记录态信息转换为自在态信息。例如自有态信息转换为记录态信息的第二个例子，生成的论文本身，不管是印刷态还是电子态，都是一个客观存在，又成为自在态信息的发生源。

第七，自有态信息之间的转换。一个最常见的可理解转换是翻译，将主体不理解的语言翻译成可理解的语言。

第八，记录态信息之间的转换。一个最常见的转换是将各种记录信息转换为二进制，转换为计算机、信息系统或网络可处理、传输、存储的对象。

信息转换是信息运动的重要组成部分，将在第2章展开讨论。

1.3 信息的一般属性

本节仅讨论所有形态信息共有的属性，共六项。特殊的属性将在本书展开的章节讨论。

第一，信息是没有质量、没有能量、没有固定形态，是与物质世界、生物界和人类社会同步演进的客观存在。从宇宙大爆炸到地球生命的诞生，100

多亿年的历史，自在态信息无处不在，无时不在，许多信息留存于宇宙空间，经历百亿年时间、弥漫于广袤的宇宙空间，存在到今天并延续到今后，为科学研究提供了信息基础。这是不证自明的性质。一个交通事故的现场，从开始到处置结束的过程，是自在态信息；路边监控录像的纪录是记录态信息，进入到计算机系统进行处理和存储的另一种形态的记录信息；交通警察目击和分析录像形成的事故现场情景是自有态信息。在这里作为分析判断这起交通事故的信息是什么？录像机和交通警察的眼睛得到了光子，光子具有能量，但光子只是事故现场实况信息的载体，而不是信息本身，在录像机中变成了电磁信号，在大脑中变成了化学物质构成的生物信号。而交通警察写下的书面记录，则是以文字为外壳、以纸张或电磁信号为载体的另一种形式。作为自在态、自有态、记录态信息交集的信息集合 F 中的信息，都可以作为信息没有质量、能量和固定形态的证据。

推论 1.1: 从自在态信息到自有态信息再到记录态信息是宇宙发展的必然规律。图 1.2 以宇宙大爆炸为起点，展示了地球文明客观演进过程，再与图 1.3 对照，这个结论也是显然的。

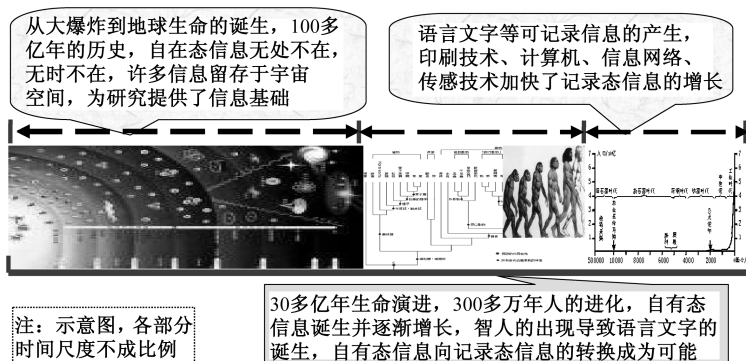


图 1.2 宇宙、生命、信息演进

图中集合 Z 、 Y 、 L 的含义与 1.2 节相同，分别表示自在态、自有态、记录态信息。只要图 1.2 为真，那么图 1.3 也为真。没有自在信息，就没有自有和记录信息，图中不包含于集合 Z 的自有和记录态信息也就不存在。

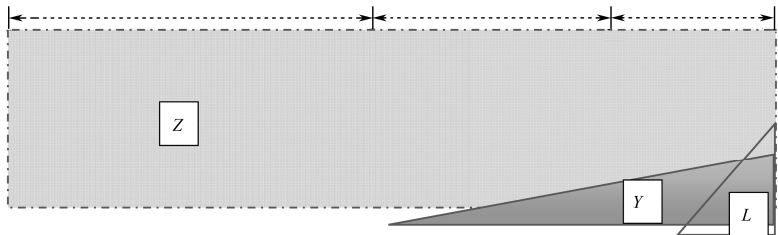


图 1.3 信息三态的历史进程

第二，信息与承载信息的载体和表示信息的外壳存在可分又不可分的统一性。三态信息，本质是含义，作为含义，信息可以与承载的载体及表示其含义的外壳分离，这在前面已经证明。但含义自身不能独立存在，作为客观存在，信息又必须具备载体与/或表示含义的外壳，即使是自用态信息，也依赖于智能体具有相应功能的物理载体。所以，作为存在的信息，使含义、载体与/或外壳的统一体，这里载体与外壳并不一定需要同时存在，有的并存，如书本，有的只有载体，如相片。原因在于外壳一定需要载体，而载体可以承载含义而单独存在。因为信息本质是含义，含义可以在三态中转换，同样含义的信息其载体是可变的，可以用不同的方式表示，所以存在信息与载体可分、又不可分这一看似矛盾而有统一的独特性质。正是这个性质，才使得信息既可以成为一类并列于物质、能源、生物体、人类社会的独立存在，又成为生命和智能演进的基础。形与义。

推论 1.2: 载体、外壳与含义可分又不可分，决定了同一含义可以存在于不同的载体或外壳中。分析信息，研究信息发生发展的规律，既要研究含义，也要研究载体和外壳。为方便起见，以后本书将与物理载体连接在一起，主要研究载体或外壳的信息称为形式信息，将只指内在含义的信息称为语义信息。

推论 1.3: 信息存在多种表示方式。同一语义信息存在于不同态的外壳中，说明表示同一语义的符号体系不唯一，表示语义信息内在关系的概念体系同样不唯一。

第三，信息在数量上既是无限的又是有限的。自在态信息是无限的，自

有态和记录态信息是有限的。

推论 1.4: 自在态信息只有变成自有或记录才可能被利用, 所以, 任何智能体及智能体集合能利用的信息总是有限的。

推论 1.5: 可用信息增长没有上限。可用信息的增长源泉是自在态信息及智能主体产生的不属于自在态的信息, 其中第一部分是无限的, 第二部分是有限的, 总合起来是无限的。

第四, 信息间存在可表示的相互关系, 语义信息间存在必然的客观关系。形式信息间存在可表示的相互关系, 图书馆的目录、数据库的倒排表、遗传基因的表达序列标签 (EST) 均是例子。语义信息间存在可表示的相互关系。人工智能的研究揭示了大量语义信息间关系表达的规律和实例。例如, 罗格著的《人工智能: 复杂问题求解的结构和策略》一书的第三部分, 以三个章节论述了这一问题⁹。语义信息间的关系是客观存在的、必然的, 因为信息是客观存在的反映, 客观存在本身具有特定的属性和运动规律, 信息增长的一个目的, 就是将这些属性和规律以记录态信息表示出来。

推论 1.6: 语义信息间的关系定义为信息结构且这一结构是可以形式化地表示的, 这样的表示可以为特定的智能体理解。

第五, 信息载体按物理与/或生命规律运动。信息的载体, 如果仅是物质形态, 而不同时又是生命形态, 则不管是什么材料或形态, 光子、声波、电磁信号、纸张等, 都是物质, 物质按物理规律运动。既是物质又具有生命特征的载体, 按生命规律运动, 遗传信息和认知信息都是如此。尽管从物理学的角度看, 遗传信息和认知信息的运动基于物理规律, 但所有的物理过程都叠加了生命运动的规律。

推论 1.7: 形式信息处理与语义信息处理遵循不同的规律。形式信息的对象, 不管是载体还是符号性外壳, 在物理空间的处理过程是形的运动或转换, 均服从物理规律。在生物体内的运动过程既遵循物理规律, 也遵循生命规律。语义信息的对象是含义, 处理对象是含义的关系, 处理目的是基于语

⁹ George F. Luger: Artificial Intelligence: Structures and Strategies for Complex Problem Solving, Fifth Edition, Pearson Education Limited 2002, 2005.

义 的表示、搜索、匹配、连接、增长、赋值、问题求解、行为决策等，需要基于语义的逻辑，及由此决定的算法。

第六，信息是智能的源泉和归宿。

如图 1.4，自在态信息从地球一诞生就弥漫于全部地球空间。35 亿年前地球开始出现生命，自有态信息随之诞生。冷热干湿、风雨雷电、大气开始为生命感知，生命为生存而应对，逐步改变了遗传基因，从简单生命像复杂生物进化，复杂生物对信息的感知和处理能力持续提升，不仅促进了遗传基因的进化，还促进了第二类自有态信息能力的提升，在亿万年的历程中，产生了记忆，产生了语言、产生了智人，产生了表达含义的符号、产生了概念，产生了文字，记录态信息诞生。智能进化走出了依赖个体、依赖人与人的口口相传，随着生产力的进步，记录态信息不断增加，人处理接受、处理信息的能力持续增强，具备该能力的人数快速增长，信息传递和处理的基础设施和能力相应提升，人类的智能水平及发展速度进入到前所未有的高度。在这个过程中，信息既是基础，又是推进器。

推论 1.8：从自在态信息经由自有态信息再到记录态信息是宇宙发展的必然规律。

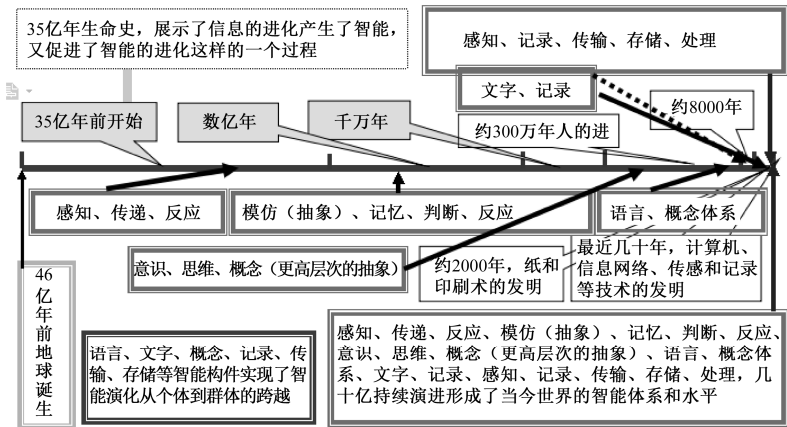


图 1.4 信息与智能进化

第2章

信息的运动

信息没有质量，没有能量，甚至没有固定的形态，但在几十亿年间持续运动和发展，这是宇宙运行中的一个奥秘。

这个奥秘植根于信息的载体、外壳、含义三位一体，按不同的规律、朝着确定的方向跨域运动。

信息的运动按不同的维度划分有不同的类型：形式信息和语义信息的运动，无主体参与的运动和有主体参与的运动等。本章从六个方面讨论信息的运动，分别是信息三态的六种转换，信息的空间移动、同态信息变换的信息处理、信息载体的积累、外壳演进过程及语义的运动。其中有许多类型的理论和实践都很成熟，或者有很多的研究成果，本书不展开分析，有兴趣的读者可阅读相关文献。

2.1 信息形态转换中的运动

信息不同形态间转换是信息发生发展的基础，信息运动的分析应从具体的转换开始。

定义 2.1：信息形态转换。信息形态转换是指一个信息集合从一种信息形态转换为另一种信息形态。这里，信息集合是指这个信息集合所包括的含义，由一种状态转换到另一种状态可以是集合中全部信息含义的转换，也可以是部分转换，即转换前后两个信息集合并不要求相等或一致。如图 2.1 所示，信息三态间的转换发生在图中相应弧线上，即三态的边界部分。

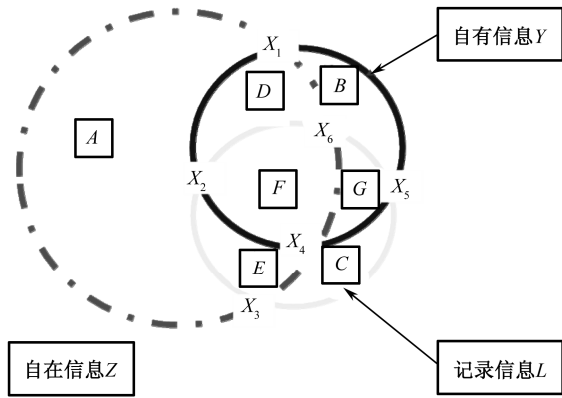


图 2.1 信息转换边界

信息转换也可以形式化地定义为：一类信息形态的一个信息集合 A 经过一个转换过程变成另一类信息形态的信息集合 \hat{A} 的过程，即 $A \rightarrow \hat{A}$ 。 A 为非空集合， \hat{A} 可以在含义上与 A 相等，也可以是 A 的一个子集，甚至是空集。其中 \rightarrow 表示实现这一次转换的转换过程。依据这样的定义，信息的三种形态转

换的组合共有六种，下面分别讨论。

2.1.1 自在态信息向自有态信息的转换

自在态信息向自有态信息的转换发生在图 2.1 中 $X_1X_2X_4$ 这条蓝色边界上，由自在态红色圈向自有态蓝色圈运动的过程。这一运动是信息发展过程中的基础环节。尽管它不是最早，也不是最广泛，最早和最广泛均由自在态信息本身的运动占有，但它走出了信息发展的第一步，无论是对于生命、智能，还是对于信息的发展本身都是最具影响力的一步。

自在态信息向自有态信息转换到现在为止主要有两种方式，一是生命体感知自在态信息，通过神经系统转换为自身可用的信息，转换的目的有三种：即时应对、生命进化、智能增长，这三种的过程将在自有态信息内部转换一节讨论；二是自动化系统（或称为智能系统而尚不具备真正意义智能的系统）为系统功能而通过传感器获取信息的运动。当具有完整的智能意义的非生物智能体产生之后，还有第三种模式。第三种模式在运动方式上与第二种模式一致，在意愿和目的上与生命体一致。

各种自在态信息向生命体的自有态信息转换运动，具体形式都是生命体自身的感知，生命体感知特定外界信息并转换为生命体可用的信息。转换的数量和质量取决于生命体感知的能力和意愿——即动力。感知的动力是生命体生存和发展的需要，感知的能力依赖于生命体自身的进化。

不同类型生命体感知的意愿和能力具有重大差别。所有生命体都有即时应对、生命进化这两类需求，具有智能的生命体包括全部三种需求。因此，自在态信息转换为自有态信息的动力和感知能力是生命体与生俱来的。触觉、听觉、视觉是生命体感知的主要渠道，但是否具有，则是生命体自身的能力。

非生物体的自动化系统或智能系统，则通过传感器，系统将传感器获得的信息传送到相应的处理器处理。

2.1.2 自在态信息向记录态信息的转换

自在态信息向记录态信息转换发生在图 2.1 中 $X_6X_2X_3$ 这条黄色线条的边界，经过一次转换过程，自在态信息转变为记录信息。

这一转换过程的产生是人类智能进化的结果，人类社会进步发展到足以不经过人的器官或大脑，直接将自在态信息转变为记录信息。这一运动的基础是技术设备，动力是记录态信息的有用性，即记录后产生的利益。记录是有成本的，只有收益大于成本，这样的转换过程才能持续，当然，这里的利益不仅指直接或间接的经济利益，还包括可能存在的其他利益，政治的、文化的、军事的、学术的等。

从运动的具体形式看，主要有两类：一类是记录与直接的系统运行连接；另一类是没有直接的后续行为。各种摄录像、录音等转换都属于前者，而人类经由认知再将认知结果转换为记录态信息的，以及作为信息系统的任务的一个前置环节的转换都属于后者。

2.1.3 自有态信息向自在态信息转换的运动

自有态信息向自在态信息转换发生在图 2.1 中 $X_1X_2X_4$ 这条蓝色线条的边界，由自有态蓝圈向自在态红圈的运动。这类运动主要表现为以下三种模式：

第一，思维与作品同含义的艺术品。雕刻、书画等艺术品，客观存在的作品所包含的含义就是艺术家大脑中的东西，在含义上是等价的，而不等价的则不属于自有态向自在态的转换。这种运动在所有的信息运动中数量最少，大量经由智能主体加工、记录变成自在态信息属于 2.1.5 节范畴。

第二，生物体生长过程中自有态信息的作用。每一个生物体都是客观存在，不管是整个生命过程还是每个时间断面的状态，都在散发出自在态信息。这些自在态信息并不全部由自有态信息决定的。尽管遗传基因及感知、反应，学习、决策、行为等过程在生命过程中具有决定性的作用，但其他外部环境对生物体的生长过程作用不能忽视。

第三，生物体内自有信息向自在态信息的转换。遗传基因既是自有态信息，生命体可以利用它复制新生命，控制生长过程，也是自在态信息，是蛋白质、核糖核酸、碱基对等生命构成部分之化学物质的结构、状态和变化。在神经系统和大脑中运动、处理和存储的信息既是自有态信息，也是自在态信息，但两者实际含义不同，作为自在态信息，内在含义是这些系统（生命体所有能传输、处理、存储信息的器官、组织或系统）的物理化学特征的变化，如视网膜感知颜色和变化及其向大脑的传输过程中不同视色素吸收不同波长的光的变化；而作为自有态信息，大脑处理的是所看到物体的颜色、形状等含义。

2.1.4 自有态信息向记录态信息的转换

自有态信息向记录态信息转换发生在图 2.1 中 $X_2X_4X_5$ 这条蓝色线条的边界，生命体或其他智能（自动化）系统将加工过的自有态信息转换为记录态信息。

智能（自动化）系统将系统内的自有信息根据一定的目的输出，不管是转变为人可以理解、阅读的文本还是输向（变换形式或不变换）其他信息系统中都属于这一运动模式。

自有态信息向记录态运动的主要部分是智能生命体（人）将接受与/或思考的结果通过一定的方式，如书写、录音、摄像、输入计算机等计算存储设备，转换成记录态信息。

这一转换模式是人类智能走向新阶段的重要里程碑，也是信息增长发展过程的重要里程碑。首先，使得人类信息交流和传承脱离了人与人直接接触的桎梏，逐步走向跨时空的群体性增长；其次，在这样的过程中，人的记忆、学习、逻辑思维能力，符号体系和概念体系的完善和发展获得了客观的基础；最后，信息本身则从最简单的直接反映（自在态信息）经由个体利用深化（自有态信息）跨入自我生长的新阶段，关于信息发展阶段，将在第 4 章详细阐述。

2.1.5 记录态信息向自在态信息的转换

记录态信息向自在态信息转换发生在图 2.1 中 $X_6X_2X_3$ 这条黄色线条的边界，由记录态黄圈向自在态红圈的运动。任何记录的结果就是自在态信息，记录信息一定是自在态信息。但是，除 2.1.2 节自有态信息向自在态信息转换表述的第一种情形外，一条客观记录，作为记录态信息和作为自在态信息不是同一个信息。例如本书，作为自在态信息，是指这本书作为客观实在的信息，如开本、页数、包装、印刷特征、封面颜色、载体特征（如纸张规格、数字存储介质及其格式）等；作为记录态信息，是指书中的内容，每一章、每一节、每个字的含义。

2.1.6 记录态信息向自有态信息的转换

记录态信息向自有态信息转换发生在图 2.1 中 $X_2X_6X_3$ 这条黄色线条的边界，由记录态蓝圈向自有态红圈的运动。记录态信息向自有态信息的转换，使得客观存在的信息成为智能体智能化程度提升和解决问题能力提升的基础原料，成为社会进步的根本动力之一，因此，是所以产生记录态信息的目的和动力，如果不能由记录态信息向自有态信息转换，记录态信息的存在就没有意义。

记录态信息向自有态信息转换是一个由智能体个体完成的学习过程，即使是一组人学习同一个记录，相互之间存在启发作用，但转换过程还是落实到具体的智能个体。对每一个具体的记录，不同个体完成的转换过程的效率是不同的，或者说智能体个体得到的信息是不同的。每个具体个人，获取记录态信息的机会是不均衡的，取决于这个人在社会中的位置和其他必要条件。因此，讨论记录态信息向自有态信息转换的数量和质量要两个方面：记录的可获得性，获取者的能力或意愿。

第一，记录信息的可获得性。可获得性首先是经济问题，对需要的记录信息是否支付得起。其次是边界问题，需要的记录信息是否有跨越边界的约束，如果存在，即不能得到。在现实中，有的记录不能跨越一个机构或一个

国家，有的甚至只能保留在一个严格保密的狭小物理空间。

第二，获取者的能力或意愿。首先，获取者对记录信息理解的能力。文本性记录信息以文字及概念作为外壳，而文字有多种体系，不同的体系需要学习或翻译；不同的文字体系或文化背景造成概念体系的差异，需要有比较调整的工具。更重要的是，相对于同专业领域的记录信息，存在获取者自有信息框架与所获得信息的水平差距，获取者的水平如果与该记录包含的信息存在较大的水平差距，所得就很少。相对于不同专业领域的记录信息，更有能否理解的问题，不能理解，一切都是零。其次是获取者的意愿问题。意愿是动力，没有主观意愿，只有外部压力，积极性就不高，效果就差；没有外部压力，就没有效果了。

2.2 信息的移动

如果说信息态之间的转换是信息构成要素中含义的运动，则信息的移动是信息构成要素中载体与/或外壳的运动。

定义 2.2：信息的移动。信息的移动是指同一形态信息的载体与/或外壳在物理空间的一点移动到另外一点。这里没有对移动的形式作约束，是指所有形式，但对信息的形态有约束，是指同态信息的移动。如果移动过程中出现了信息形态的变化，属于信息转换的范畴。尽管移动的目的是载体或外壳中所承载的含义，但空间的移动必然要求载体与/或外壳作为依托，所以本节定义的一定为形式信息的空间移动。移动是信息运动最直观的形式。有些移动符合这一定义，如上节关于信息态间转换过程中存在外壳或载体的移动，以及在信息积累和处理等章节中涉及的物理空间移动，但本节不讨论。

2.2.1 移动的类型

信息的移动也可以形式化地定义为：一类信息形态的任意信息集合 B 从物理位置 L_1 经由一次传输过程，成为存在于一个新的物理位置 L_2 的信息集合

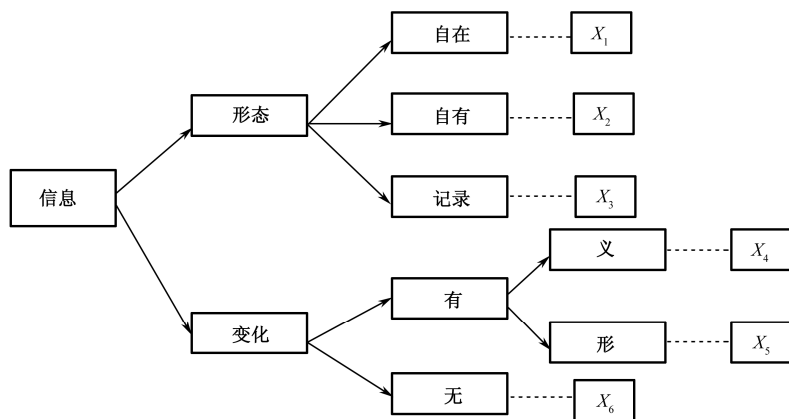
\hat{B} 的过程, 即 $L_1 \rightarrow L_2, B \rightarrow \hat{B}$ 。其中 \rightarrow 表示实现这一次移动的传输过程, 及在这个过程中信息集合 B 的变化, 当然也包括没有变化。这个定义说明, 信息的移动形式看是载体与外壳的移动, 而本质上还是信息集合所包含的内容的移动, 但移动过程遵循物理或生命规律, 而不是信息含义的规律。本节讨论移动的要素和类型。

第一, 信息移动的要素。一个信息集合从空间一个点移动到另一个点, 有两个要素, 一是信息集合本身, 二是通道。

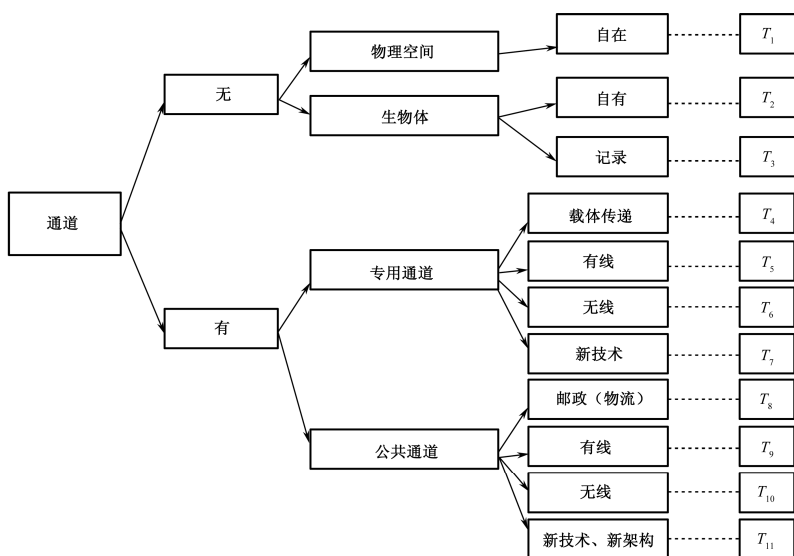
第二, 信息移动的类型。信息为什么移动? 就信息而言是自身更好的显示, 就使用者而言是为了更好的使用。多样性的移动就是显示和各种形式用的需求驱动的结果, 但具体的移动与显示和用的关系, 则存在不同的情景, 呈现出不同的移动模式。两个要素的细分和组合, 展示出信息移动的丰富多样和规律, 图 2.2 给出了基本的信息移动类型。

图 2.2 的 (a) 和 (b) 两个部分列举了信息移动的基本类型。其中 $X_1 \sim X_6$ 是关于所移动的信息集合的分类, $T_1 \sim T_{11}$ 则是关于移动借助的通道的分类, 其中二级分类“有”与“无”是指专用于信息传输的通道还是借用了一般性的运输通道, 也是为了区分“有”的下一级分类而定的名称。实际的信息移动过程则是两种分类的组合形态。

$X_1 \sim X_3$ 是自在态、自有态、记录态信息移动的分类。 X_1 是指自在态信息的移动。自在态信息的移动依据物理规律, 主要是客观世界的存在在光、声、辐射等物理作用下, 从所反映的存在体向可达物理空间的移动。这一移动对信息可感知具有重大意义, 也为自在态信息向自有态、记录态信息发展提供了有利条件。 X_2 是指自有态信息的移动。自有态信息的移动遵循智能生物体活动规律, 生活空间的迁移、生活或工作需求的空间移动。 X_3 是指记录态信息的移动, 记录态信息产生的原因就是为跨时空的保存和交流, 记录态信息的移动是信息移动的主要形态, 有时候, 信息移动就特指记录态信息的移动, 图 2.2 (b) 部分中的主要模式都是用于记录态信息移动的。要区分的一点是, 语音信息移动, 若是在人群中口口相传, 属于自有态信息移动, 若是经由电信或其他相同功能的信息网络传输, 则属于记录态信息移动, 因为进入这样的系统, 就从自有态转换成了记录态, 到接收端是一次信息外壳的再转换。



(a)



(b)

图 2.2 要素约束下的信息移动类型

$X_4 \sim X_6$ 是对于一次具体的移动过程中，作为移动对象的信息集合是否产生了变化，产生的是什么变化。 X_4 是指一次具体的信息移动中，信息集合包

含的含义有了变化，变化包括增或减。增的概率远低于减的概率。发生增减的原因很多。例如在互联网传输时丢包，人在交流时失真等。 X_5 是指一次具体的信息移动中，信息集合的载体或外壳发生了变化。这种变化有时是必要的，例如为信息传输通道的技术要求，要进行编解码。有的时候是由于传送通道失真，这种失真在传输过程的各个环节度可能发生。 X_6 是指在信息传送过程中没有发生含义或载体的变化或丢失。绝大部分信息移动过程都能保持信息集合的完整一致，无误差从出发地送到目的地，在复杂的信息传送通道中，一般通过纠错的模式保证正确性，系统越复杂，信息传送不失真的要求越高，纠错功能就越强。

$T_1 \sim T_3$ 是指自在态、自有态、记录态信息不经由各类建设起来的信息传递通道的自身移动。其中 T_1 是自在态信息在可达空间的移动，与 X_1 组合在一起，形成完整的自在态信息移动过程。 T_2 是指无通道式自有态信息移动。自有态信息有两类存在状态，一类在生物体内，一类在私有的信息系统内，这里的私有是与面向公共服务的信息系统相对而言。在信息系统内的自有信息传递必然经由专门的通道，因此从通道看自由态信息的移动，特指生物体。生物体两类自有态信息中，遗传基因的移动在生物体内部。因此，自有态信息移动就是指人或其他智能生物体在移动中带来的信息移动。 T_3 则是记录态不经特定管道的移动。典型例子是一个学者带着学术资料，参加一个学术会议形成的一次记录态信息移动，记录态信息从一个地方移动到了另一个地方，而没有经过任何商业化或专有的信息传输渠道。

$T_4 \sim T_7$ 与 $T_8 \sim T_{11}$ 所使用通道的技术类型几乎一样，传输的也都是记录态信息，不同的仅是通道归谁所有。专用通道归特定的机构拥有，仅为该机构的信息移动服务，公用通道则面向全社会开放，随着技术的进步，专用通道与公用通道的衔接更加密切，专用通道承载的信息移动在全社会信息移动中的比例呈持续下降的趋势，而且除开少数特例，如保密、规模过小等，几乎所有的专用通道利用共用通道的资源与/或服务。 T_4 是指记录态信息载体（不是外壳）的直接送达模式，如磁带、磁盘、书籍等的内部交换。 T_5 和 T_6 是指机构内的有线及无线信息传输网络，由于信息技术的进步，这两种模式正在融合为一个统一的系统。这里所指的机构包括各类不同的企业、学校、医院、

科研、非政府组织、政府及其他公营机构等。由于机构的规模、性质差异大,所以这些机构所拥有的信息传输网络的规模、性能、架构体系等存在巨大差异。既有像阿里、腾讯、百度、沃尔玛、IBM、CISCO 那样遍布全球、性能先进、有线无线一体的企业网,也有小企业、小服务机构的小规模有线与/或无线的局域网,还有依托共用信息网络构建的私有虚拟网。 T_7 是指已经出现或将要出现的新技术构成的信息网络,可见光、量子等新的信息传输介质可能会改变机构信息网络的架构。

$T_8 \sim T_{11}$ 是服务于信息移动的共用通道,在社会信息传输中扮演最重要的角色。 T_8 是指公共的以信息载体为对象的信息移动通道,主要是邮政系统、物流系统中运输信息载体的部分。特殊形式的传递,如古代的烽火报警、驿站速报,现代的旗语、专人专递等属于这一类型。 T_9 、 T_{10} 包含了今天所有电磁形式的信息移动共用系统,以有线与/或无线方式组成的电信网、互联网、物联网、广播电视网、电力线构成的信息网络、卫星通信网、对地观察系统等。 T_{11} 是指在通信技术的推动下,共用信息传输通道的变革。这些变革包括新技术容纳进已有的信息网络中或新技术推动下,颠覆已有的信息传输技术体系架构,构成新的体系架构,如天地一体、无缝连接的信息传输网。

2.2.2 信息移动过程的环节与分解

根据定义,信息移动的起点是信息提供方根据信息需求方(包括提供方自己,因为显示拥有的全部或部分信息是许多提供方的意愿)的要求,经由信息移动通道到达信息接收方的过程,而满足需求方接受和通道传输的要求,还有三种不同的转换。因此,一般意义上的信息移动由六个部分组成:信息提供方、信息需求方、信息移动通道、信息可提供性和可接收性的处置。

在图 2.3 中,信息提供方和需求方分别给出了四个示意性图标,分别表示智能生物体(地球文明而言主要是人)、传感器、数字设备(机器)及私有

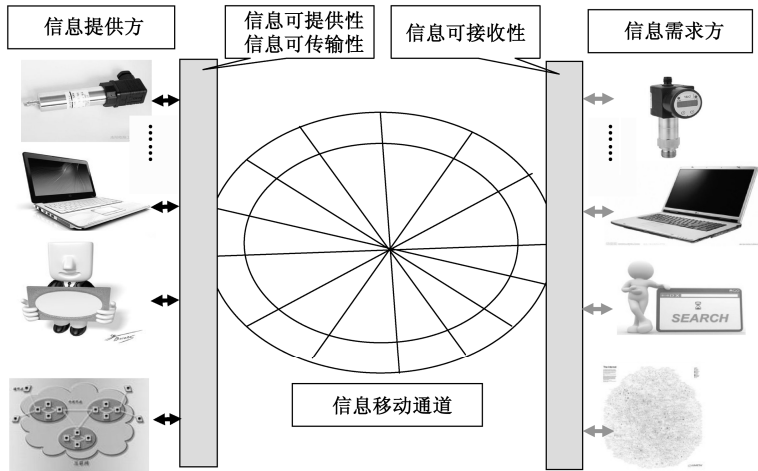


图 2.3 信息移动过程
(图片源自互联网)

信息系统、公共信息系统这四种主要的信息提供和接收方式，实际上是泛指所有的信息提供方。这四类的每一类都是一个庞大的家族，实际的提供和需求比这四类要复杂得多，有更多的组合态，如智能生物体与其他三类的组合，传感器与私有或公有信息系统的组合，2.2.1 节中讨论的无通道信息移动类型也没有包括在内。图中的通道泛指一切承载信息移动的通道。而提供方、需求方两者之间以及这两者与通道的传输要求之间以及如果存在需要载体或外壳的转换，则经过两侧的可提供性、可传输性和可接收性的转换。

设信息提供方为 S ，信息接收方为 R ，通道为 C ，可提供转换为 ST ，传输转换为 CT ，可接收转换为 RT ，信息移动各环节形成如图 2.4 的关系。

如图 2.4 所示，不计间接组合关系，六个环节两两之间存在着 30 种直接关系。而具体信息移动的关系大部分是组合关系，如 $S-C-R$ ， $S-ST-CT-C-RT-R$ 等最重要的过程。

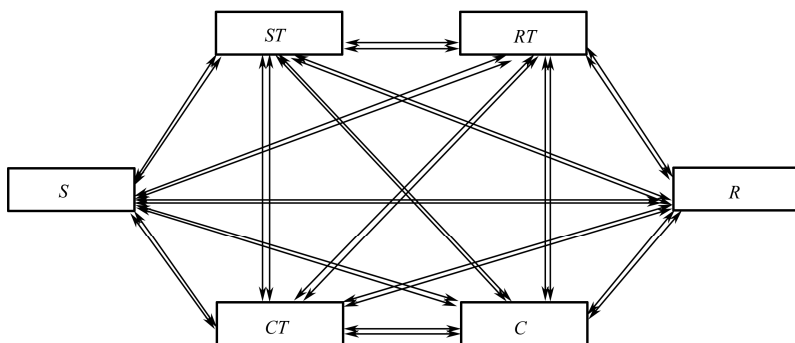


图 2.4 信息移动各环节的关系

在信息移动过程中，对 S 、 C 、 R 的关注比较多，特别是 C ，但对 ST 、 CT 、 RT 的分析比较少。 ST 是可提供性转换，这一转换是许多具体信息移动所不可或缺的环节。一是关于价值的审核，公开还是机密、免费还是收费，收费多少合理，经过这一过程，信息集合才具备可移动性。二是关于形式的转换，发布性的要做适合潜在对象偏好的转换，按需求的则要符合接收方的要求。有一种转换可能更加复杂，就是跨信息域信息移动中的可提供与可接收转换。信息域是指使用相同的符号及概念体系的信息流通圈，也就是在这个域中，信息交流不存在符号与/或概念的障碍。跨信息域移动是指信息移动跨越两个存在不同符号与/或概念体系的信息域。这是信息移动过程中的一个重要环节，如果不能成功转换，信息移动没有意义。中文甲骨文到简体中文，繁简体转换，中文到其他文字的转换，以及任何不同语言文字体系的转换，都属于此。这些转换主体是符号体系的转换，在一些概念和符号体系不完善的文字间的转换，涉及一定的概念体系转换。如果出现地球外文明信息与地球文明信息的转换，可以预见是符号体系到概念体系的全面转换。符号体系转换是概念体系转换的前提，概念体系转换是信息增长的重要方式。从信息提供的显示度目的出发，还有一类重要的 ST ，就是将提供方准备放到公共信息提供场所（互联网、平面媒体等）的信息集合要做适合潜在接收者偏好的格式性转换，包括文字的风格、表现的风格等。

CT 是提供方或接收方通过传输通道时进行的适合性转换,不涉及传输过程中可能发生的转换。所谓适合性,就是提供的有效性、传输的有效性、接收的有效性,这也是信息移动追求的目标。满足传输有效性的转换在通信科学领域做了大量的讨论,申农及诺斯、科捷尔尼科夫、乌特沃德等在信道编码定理、通过率失真函数、噪声理论、最佳接受理论、雷达监测理论等方面为 *CT* 建立了理论基础。在一般的信息移动过程中,还有大量在上述研究之外的 *CT* 环节。不同信息集合间冗余的处理,如属于不同个体,内容相同的图片或视频,如何既能实现传输的最小冗余,又能保证接收信息集合的完整。再如,同一时间的相同或不不同的需求的转换,如何实现最小传输和最佳分配。

RT 对许多以传感设施为提供方的信息移动变成十分重要,因为一般的传感器获取的信息没有可读性,需要经过不同复杂程度的转换,才能成为系统可处理或人可理解的信息。卫星遥感、太空观察、基本粒子运动观察、多普勒雷达对大气层的扫描、脑细胞活动的观察等,没有 *RT*,就没有感知结果的利用,也就是一次无效的信息转换(从自在信息转换为记录信息),也是一次无效的信息移动。在绝大部分企业的自动化或智能系统中,实现这个转换是系统的重要功能,感知设备对温度、湿度、压力、空气洁净度、材料内部损伤、有害气体、杂质等的感知结果需要前置的转换功能,变成系统可处理、判断的信息,在很多系统,还有世界的约束,即在规定的时间内完成转换,获取及传递的信息才是有效的,否则就没有意义。

2.2.3 信息移动的约束

从图 2.3 和图 2.4 信息移动的要素和环节的分析,有效信息移动过程存在一系列约束,*S*、*C*、*R*、*ST*、*CT*、*RT* 各自及其相互关系也在发展中,信息移动模式经历了多次质的变革,还将继续演进。

信息可提供约束如图 2.5 所示,主要有两个方面 10 个因素。第一个方面是需要的信息不存在,第二个方面是信息虽然存在但有的不能提供、有的不能利用,具体可分解成 $S_1 \sim S_{10}$ 的原因。 S_1 是指需求的信息不存在。信息三态转换中的障碍、还不存在转换为记录态信息的前提以及已经存在而物理性毁

损,如秦始皇的“焚书坑儒”。信息不存在既可以看作信息移动的起点问题,也是信息增长的起点问题,是信息发展所需要解决的根本问题,将在本书第4章展开讨论。 S_2 是指一个基本的信息,已经存在于特定的记录信息集合中,也可以提供,但由于地点或数量的原因,只能满足部分需求,还有部分需求自给定时间内不能满足,这样的情形在互联网诞生之前广泛存在,互联网普及之后有所改善,还没有完全解决,在一定的意义上,这是必然现象。 S_3 是指从信息提供方到信息接收方之间物理不可达。在记录信息产生到完善的邮政、信息网络形成之前的数千年时间内,物理不可达是信息移动的主要制约因素。即使到了20世纪末,如果计算全世界所有应该得到一个具体有用信息集合的满足率,相当部分还不足1%,即使到了互联网十分普及的今天,满足率不足1%的例子依然大量存在。 S_4 是指潜在的信息接收方如果得到一个特定的信息集合后,可能与提供方的利益发生冲突,因此采取选择性提供的方式,许多具有商业意义的信息,也包括学术性的信息,存在此类约束。 S_5 是指法律法规和制度规定的信息提供约束。如国家机密、个人隐私、对政治社会产生重大影响的敏感信息,都需依据相应的法律法规和制度执行,不能以需求为准。 S_6 是指信息移动的文化约束,特定的信息在一定的社会域:国家的、地区的、宗教的、意识形态的等原因,不能按需求移动。 S_7 是指提供信息的载体不适合接收方的使用能力或超出了需求方能力。数字或模拟的信息都可能存在这个问题。看起来这是一个信息转换的问题,但实际上要复杂得多。 S_8 是指提供信息集合表示信息的外壳不能为接收方理解。在语言文字多样化的现实世界,这是一个制约信息移动的重要原因,如果一个人或一个系统,只能理解利用一种主要语言的话,意味着大部分记录信息不能利用,也不会提出需求,信息移动就失去了动力。 S_9 是指信息的含义接收方不能理解。这与 S_1 、 S_3 一样是信息移动的关键障碍。在人类记录信息形成以来的历史长河中,接收方的含义理解能力,不仅是受教育程度,始终是文明和智能进化的主要障碍。 S_{10} 是指提供到接收之间移动过程预期或实际超过了信息发挥作用的时间要求,导致了不移动或无效的移动。

在上述讨论中没有论及地外文明与地球文明间的信息移动,如果发生,约束相同的。可提供和可接收是一个事物的两面,接收方分析是提供方分析

的反向，所以不重复讨论（见图 2.5）。

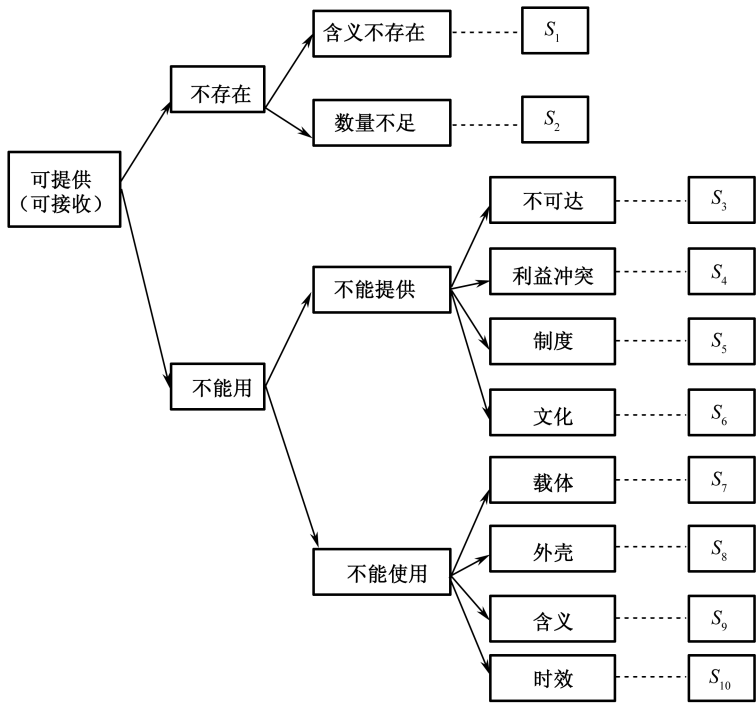


图 2.5 信息可提供性（可接收性）约束

通道对信息移动的约束是最直接的，是决定可达性的主要因素。人类不断追求信息移动渠道的畅通。通道成为制约因素，或跟不上信息移动的需求首先是技术原因。从利用声音大声呼叫，到利用可见光烽火传递，再到利用畜力驿站传送，是工业革命之前的主要方式。工业革命之后首先建立了全国性甚至一定程度上全球性的邮政系统，其次是利用了铁路、公路、飞机的传输渠道，最后，数字化信息的传输从电报、电话、计算机网到互联网，通道技术不断进步，形式更趋多元，单位时间传输的信息呈指数增长。但是，应该看到，在通道日益进步的同时，对信息移动的需求以更快的速度增长。从绝对量看，今天不能满足的信息移动需求，比历史上任何时候都要多。一种

新的技术或形式，激发了没有这种技术和形式之前没有产生的需求，而且，随着经济社会的发展，提供方和接收方的能力也在不断提升，更多的提供和需求意愿导致了信息移动缺口的增加。第二是经济原因。一个国家或地区社会信息传输通道的发展水平一般与该地的经济发展水平呈正相关。同时，信息传输通道建设需要大量的资金，每一次新的技术带来新一轮投资需求，资金是否充裕，直接影响到通道的性能和质量。第三是社会问题。信息传输通道是社会公共基础设施，也是商业运行的平台，不同的认识和体制影响着公共资源和私营部门资源的配置。即使是经济发展水平相近的国家和地区，其发展的速度也有很大的不同。

提供、通道、接收等转换环节制约信息移动发展的因素与通道基本相同，主要是技术、经济和社会。在不同的环节，三者影响程度有所不同。在提供转换中，技术和经济因素占主导地位；在通道转换和接收转换中，经济因素影响效果最突出；而社会因素在提供和接收的某些部分，具有决定性作用。

2.2.4 信息移动的演进

信息移动的演进就是信息移动能力不断提升的历程。根据对 2.2.3 的分析，可以把信息移动能力定义为：

定义 2.3：信息移动能力。信息移动空间（ M ）及给定时间（ t_i ）内的信息移动能力表示为：

$$F(M_{it}) = (S+R) \times C \times T \quad (2.1)$$

其中， S 为提供方集合，有两个主要组成部分， S_p 和 S_o 。 S_p 指提供方所有主体的集合， S_o 指所有可提供的载体与/或存储载体的系统。 R 为接收方集合，也有两个主要组成部分， R_p 和 R_o 。 R_p 指接收方所有主体的集合， R_o 指所有接收的载体与/或存储载体的系统。 C 指所有通道的集合。 T 指所有转换的集合。 t_i 指可定义的时间段。#

这样，我们有：

$$\text{提供能力 } SF = S_p + S_o = \sum_{i=1}^p S_{pi} + \sum_{i=1}^s S_{oi} \quad (2.2)$$

$$\text{接收能力 } RF = R_p + R_o = \sum_{i=1}^u R_{pi} + \sum_{i=1}^r R_{oi} \quad (2.3)$$

$$\text{通道能力 } CF = \sum_{i=1}^n C_i \quad (2.4)$$

$$\text{转换能力 } TF = CT + ST + RT = \sum_{i=1}^m ct + \sum_{i=1}^l st + \sum_{i=1}^q rt \quad (2.5)$$

这些能力均以信息的载体或外壳为基础，对信息的转换或移动，更重要的是在含义层面的能力。对含义层次的满足度、通过率、转换率等评价指标的讨论见本书第 5 章。

在信息移动发展历程中，无论是主体还是载体，无论是通道还是转换，都发生了多次质的变化，成为信息增长的重要动力，在第 4 章详细讨论。

2.3 信息处理

转换、移动、积累等各类信息运动均需要不同程度的处理，处理是信息运动的基础。

2.3.1 信息处理的定义

定义 2.4：信息处理。本书所指信息处理是：根据特定的指令，对一个确定的信息集合利用相应的工具或方法进行变换，达到确定目标的过程。这个定义包含五个要点，一是目的性，每一项信息处理是根据特定主体发出的指令进行。二是范围在一个特定的信息集合。三是利用相应的平台，使用特定的方法。四是具有明确的目标。五是一个过程，不是一个片段。这个定义没有将信息处理约定在一个具体的领域，或具体的处理平台、工具或方法，囊括了全部信息处理类型，但在各个方面赋予规定性要求，强调了目的性。

2.3.2 信息处理的类型

按照上述定义,信息处理包括的范围很广,类型很多,用不同的分类标准划分则会产生不同的子类。本书按照处理对信息集合所做的变化为标准,信息处理的类型如图 2.6 所示。

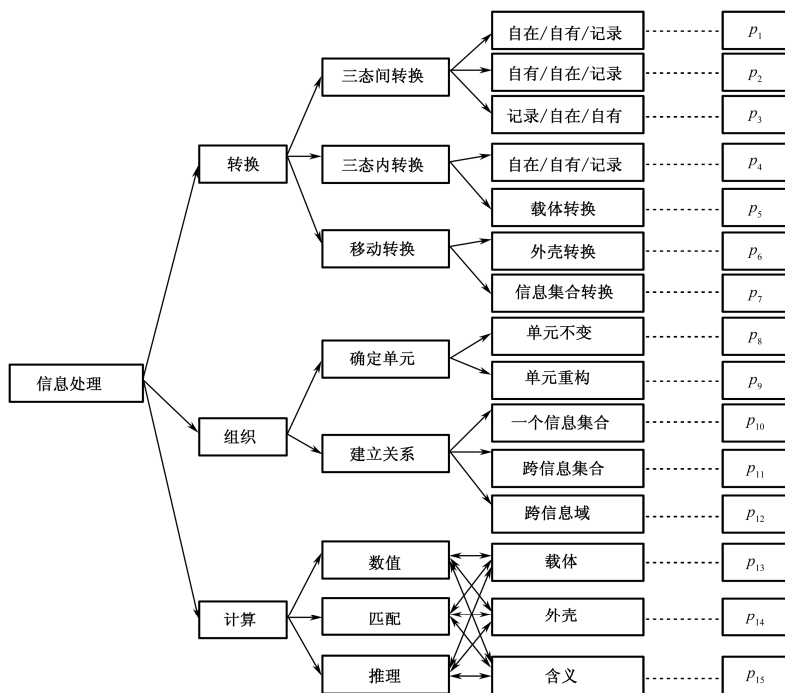


图 2.6 信息处理分类

如图 2.6 所示,信息处理按信息运动的分析框架 3 个一级类,8 个二级类,21 个三级类。如果组织与计算两个一级类区分自在、基因、认知、记录的不同处理方式,则有更多的 3 级类。

2.3.3 转换类信息处理

转换处理是指信息三态之间和内部一个信息集合到另一个信息集合的转换，形态不同，集合中信息相同。从 p_1 到 p_3 的转换在本章第二节中已经分析，本节不重复。三态的内部转换 (p_4) 在信息处理中有很特殊的位置，是理解信息处理统一性的钥匙。

自有态信息内部的转换处理发生在图 2.1 的蓝色圆圈中。在三种形态信息内部转换中，自有态信息最为复杂，包括知识性、行为性、生命性、群体性四类，使用的工具性质也存在重大差别，但实现的功能、转换处理的要素是一致的。一种是知识性内部转换处理。记录态信息转换为自有态信息，也是一种知识性内部转换处理，这里不再讨论。另一种是知识性内部处理，是人脑将大脑中已有的知识和感知到的自在态信息或记录性信息连接在一起，进行学习、分析、决策、行为等目的而使得大脑知识性信息不断增长的信息处理。这种处理的动力来自人对未知探索和信息追求的本能。

在经历一件事时，同时兴奋的脑细胞之间会建立永久的连接。

接下来，这些连接会让同一个脑细胞的网络再次兴奋，也就形成了我们的记忆。

新的经历会让脑细胞网络发展出更多的连接，让记忆更加丰富，帮助我们学习，但有时也会歪曲记忆的内容，产生错误记忆。

因为知识来自联系，所以学习的最佳策略是找出不同主题间有意义的联系。

框 2.1 认知信息的转换源自联想、记忆和学习，Anthony J. Greene 著，周林文译，
大脑与认知，电子工业出版社，2012 年，第 192 页。

行为性内部转换处理是指智能体（人或智能系统、自动化系统）针对当前场景获得信息，与 / 不与已有自有信息整合，经过处理，即时作出判断和行为。这类信息转换处理的动力来自人对外界环境的反应及对完成任务的需

求。这样的信息处理使人拥有的自有态信息质量不断提升、数量不断增加，吸收新的信息、判断和处理问题的能力不断提升。

生命性内部转换处理是指遗传信息在特定生命存续期间的持续作用，管理和控制着细胞、组织、器官、系统的生长、衰老。如框 2.2 所摘录的基因转录机制。

转录有三个阶段，起始、延伸和终止，以下是细菌转录的三个阶段。

1. 起始 (Initiation) 首先，酶识别一段叫启动子 (Promoter) 的区域，启动子就位于基因的“上游”。聚合酶紧紧地结合在启动子上并引起启动子内两条 DNA 链局部至少有 12pb 解链或分离。接着聚合酶开始合成 RNA 链，其作用底物是 4 种核糖核苷三磷酸：ATP、GTP、CTP 和 UTP。起始底物通常是嘌呤核苷酸。当第一个核苷酸到位后，聚合酶会把第二个核苷酸与第一个连在一起，形成 RNA 链中的初始硫酸二酯键。在聚合酶离启动子开始延伸之前会连上几个核苷酸。

2. 延伸 (Elongation) 在转录的延伸阶段，RNA 聚合酶指导核糖核苷酸连续从 5'→3' 方向加入到延伸的 RNA 链上。此时，RNA 聚合酶沿 DNA 模板移动。这一解链区使 DNA 模板的碱基逐个暴露出来，并与新加入的核苷酸配对。转录机器通过后，两条 DNA 链又彼此相互缠绕，重新形成双螺旋。

3. 终止 (Termination) 如同启动子作为转录的起始信号一样，基因末端叫做终止子 (Terminator) 的区域发出终止信号。终止子与 RNA 聚合酶联合作用，松开 RNA 产物与 DNA 模板的结合，结果使 RNA 脱离 RNA 聚合酶和 DNA，从而停止转录。

框 2.2 细菌基因转录过程¹

¹ 分子生物学 (原书第五版), [美]Robert F. Weaver 著, 郑用琰、马纪、李玉花、罗杰等译, 科学出版社, 2013 年, 第 36-37 页。

群体性内部信息转换处理是指一个智能群体的个体间未经记录信息中介的群体自有态信息交换。在文字产生之前、语言（包括肢体语言）产生之后这段百万年计的人类历史过程中，这种自有态信息内部群体性转换是人类信息交流和发展的主要模式。文字产生之后，即使到网络、移动智能终端等信息传输、记录的设施十分完备的今天，人与人之间的语言及肢体信息交流依然承担重要的社会功能。在没有文字的智能生物群体间，这类信息转换处理是这些群体维持生物群体生存与发展的必要功能。文字产生后，群体范畴不断扩大，一个社会内部的信息传递成为社会发展的重要组成部分。框 2.3 是一个关于社会信息传播的还在争议中的命题：迷因。

迷因（Meme）也称为米姆，弥、迷米、弥母以及模因等，是文化信息传承时的单位。这个词是在 1976 年，由理查德·道金斯在《自私的基因》一书中所创造，将文化传承的过程，以生物学中的演化规则来作模拟。

迷因类似作为遗传因子的基因，为文化的遗传因子，也经由复制（模仿）、变异与选择的过程而演化。举例而言，某个人类大脑中的观念（迷因），经由模仿或是学习复制到不同人的大脑中。而经过复制的观念并不会与原来观念完全相同，因此产生变异。这些相似但是有所不同的观念，则在散布时互相竞争，因此出现类似天择的现象。

框 2.3 迷因的讨论²

记录态信息内部的转换处理发生在图 2.1 的黄色圆圈中，是指不同记录态形式的机械性转换，所谓机械性转换，就是含义原封不动的转换，没有经由智能体或智能系统对记录的内容进行修改。记录态信息的内部运动，动力来自处理需求，是使特定信息集合便于利用。扫描仪将模拟信息的数字化就是典型例子。在自动化系统中，存在大量的模数 / 数模转换。在一定的意义上，印刷也是记录态信息内部转换的一种方式，计算机信息系统的复制同印

² 见百度百科，迷因学词条。

刷具有相同性质。信息系统输出，不管是输向系统，还是通过打印机输出，都是记录态信息的内部运动。这种方式与后述组织和计算两类信息处理有较多重合，后面再详细讨论。

自在态信息的内部转换处理发生在图 2.1 的红色圆圈中，所有自然界发生的镜像现象，如海市蜃楼、水面倒影、镜面反射、量子纠缠等均属于此。这是基于物理规律的自然现象，是转换体系中最简单的处理系统，但实现了信息转换、移动的功能。

$p_5 \sim p_7$ 的转换处理是指同态信息的转换，为信息移动或其他目的把一个信息集合转换为同态另一种表达方式（载体或外壳）的处理。在讨论信息转换和移动中对这种运动过程已经分析过，这里不再重复，也进一步说明了信息处理的基础性。

2.3.4 组织类信息处理

组织类信息处理（ $p_8 \sim p_{12}$ ）是指对一个给定范围内的信息集合按一定的规则与/或目的进行处理，使集合中的信息更加适合利用的需求。组织类信息处理只发生在可以历时积累的信息形态中，主要是记录、基因、认知类信息，而基因、认知信息的组织处理由生物体内在功能承担。记录信息的组织形式会影响到认知信息的组织，这是人类的学习功能，将在下本《智能原理》一书中讨论。这类信息处理存在的历史很长，不断从简单走向复杂，从不大的信息集合到极其庞大的信息集合；从几千年前对王室保有的档案、文献以目录为基础的处理到今天广泛存在于各类组织机构，甚至个人的书目、分类、主题，数据库、搜索引擎中的组织类处理；从基于印刷的书本、数据库的字段到人工智能中的语义单元；从几十、几百个单元到百万、十亿、万亿甚至更多的单元；从一个信息单元对应一个或几个关系，到揭示该信息单元的全部需要揭示的关系；信息组织成为信息处理的一个重要组成部分。

组织类信息处理操作的对象有两个：单元和关系。对单元的操作有两类：不变与重构，而重构主要为分解、组合、衍生。对关系的操作就是在不同的信息集合确定关系，操作的对象是单元和关系。

信息的组织管理，首先是确定单元。随着记录信息的多样化及应用需求的深化，特别是感知和记录信息的非结构特征，确定单元成为逻辑性很强（在本书第3章讨论）、工作量很大的基础性信息处理工作。确定单元的依据是需求，不同的需求对相同的信息集合单元的分解或组合具有明显的不同。一个图书馆、资料室，如果以借阅为目的，单元就是最小物理单元，如一本书、一份报、一张光盘。如果以研究为目的，就需要更小的单元。对于任何信息集合，逻辑最小单元是最小语义单元，对于应用，最小语义单元是相对的。单元需要有标识，主题词、分类号、书刊名都可以是标识。恰当的标识是信息组织管理质量和利用的一个核心环节，是信息结构的基础。其次是确定关系。关系是指对象信息集合中信息单元之间存在的一种联系，这种联系可能是语义的也可能是形式的，还有介于语义和形式之间，如人名或书名按字母顺序排列的关系。按照不同的原则，信息单元间存在的关系和信息单元在关系中的位置是变化的。单元和关系的确定有两个基本出发点，一是基于信息集合内在的语义逻辑，二是基于应用目的。不管是基于语义逻辑还是基于应用需求，在认识和处理的发展过程中，已确定单元和关系依然会持续变化。单元和关系是信息结构的核心内容，将在第3章重点讨论。

在基因及认知信息的组织处理中，单元和关系的确定及组织、重组是由生物体处理信息的功能所决定的，而处理功能对单元和关系的处理能力是生命长期进化的结果，详细的讨论将在下一本书——《智能原理》中论述。

2.3.5 计算类信息处理

计算类信息处理（ $p_{13} \sim p_{15}$ ）是指在一个给定范围内的信息集合中，针对特定目的、利用特定的工具和方法，进行问题求解的信息处理过程。在记录、基因、认知类信息形态中都存在计算类信息处理。记录类信息的计算处理从分类看与认知类信息处理几乎一致，但工具和方法不同，对认知类信息的计算处理也将在《智能原理》一书中论述。基因类信息的计算处理与其转换过程高度重合，在本书第3章和第4章进一步讨论，本书讨论的重点是记录类计算信息处理。

从目的看，信息处理主要围绕三个目的开展，一是学习，二是行为，三是研究。图 2.6 显示的记录信息计算处理分类是一个 3+3 的组合，一共 9 个小类。实际上，计算类处理还有更多的维度需要细分：处理的对象有载体、外壳及含义，处理的工具有人工、机械、计算机，围绕数值、匹配、推理又有大量的计算方法。所以，计算处理是一个由庞大子类构成的集合。下面从学习、行为、研究三个目的为主线，作一个简单分析。

学习型信息计算。学习型信息计算是指围绕智能主体的学习目的，配合智能主体的学习过程而进行的计算。学习型计算可以分成两大类，一是计算机系统，如人工智能，框 2.4 简要说明了 IBM 沃森的深度学习（deep learning）系统。二是辅助人的学习的计算。从掰手指、用算盘到今天的各种计算机辅助学习功能。

沃森由 90 台 IBM 服务器、360 个计算机芯片驱动组成，是一个有 10 台普通冰箱那么大的计算机系统。这些服务器采用 Linux 操作系统。IBM 为沃森配置的处理器是 Power 7 系列处理器，这是当前 RISC（精简指令集计算机）架构中最强的处理器。它采用 45 纳米工艺打造，拥有八个核心、32 个线程，主频最高可达 4.1GHz，其二级缓存更是达到了 32MB。存储了大量图书、新闻和电影剧本资料、辞海、文选和《世界图书百科全书》（World Book Encyclopedia）等数百万份资料。每当读完问题的提示后，“沃森”就在不到三秒钟的时间里对自己的数据库“挖地三尺”，在长达 2 亿页的漫漫资料里展开搜索。（果壳网：寻找 Watson）

沃森是基于 IBM “DeepQA”（深度开放域问答系统工程）技术开发的。作为“沃森”超级电脑基础的 DeepQA 技术可以读取数百万页文本数据，利用深度自然语言处理技术产生候选答案，根据诸多不同尺度评估那些问题。IBM 研发团队为“沃森”开发的 100 多套算法可以在 3 秒内解析问题，检索数百万条信息然后再筛选还原成“答案”输出成人类语言。每一种算法都有其专门的功能。其中一种算法被称为“嵌套分解”算法，它可以将线索分解成两个不同

的搜索功能。这听起来比较晦涩但却很实用。例如：问“在（哥伦比亚广播公司）《60 分钟》节目首次播出时，当时的美国总统是谁？”沃森首先要理解“首次播出”是什么意思，以及与“首次播出”相关的日期；其次，它必须要弄清楚具体的《60 分钟》节目首次播出的日期；然后它才能搜索到当时的美国总统是谁。

框 2.4 IBM 沃森深度学习的例子³

行为型信息计算。行为型信息计算是指围绕智能主体的当前的行为需求，配合智能主体的行为过程而进行的计算，智能体是指生物智能体和非生物智能体。行为型计算可以分成三类，一是事务，二是生活，三是娱乐。事务是指一切智能体承担的业务性事项，各类工作、学习、研究都属于这个类型。各类机构中的计算机信息系统所做的计算大体上属于这个范畴。生活是指一切工作之外为生活所需发生的行为。衣食住行、健康医疗都属于这个类别。家庭服务机器人、实时健康信息感知等均属于这类处理。娱乐是指智能体个体在生存之外的各类娱乐、消遣性行为。娱乐计算一般承载主体是智能生物体，特指人，娱乐计算的典型是玩网络游戏。应该区分的是，家庭服务机器人、实时健康信息感知、网络游戏等计算其功能实现过程分别属于生活和娱乐计算，而功能的形成及服务提供等计算属于事务计算。

研究型信息计算。研究型信息计算是指智能主体为发现以前没有发现的规律，发明以前尚不存在的技术或产品所进行的计算。各类机构或个体利用计算机和网络系统的计算功能，为发现和发明所做的计算均属于此类。

学习型、行为型、研究型信息计算在实践中既可以区分，但又存在大量交叉。有时候目的是学习，但结果又发生在行为和研究过程中，相对于行为和研究也一样。

2.3.6 信息处理的要素和体系架构

分析上述各类信息处理，可以抽象出相似的要素及体系架构。信息处理

³ <http://discover.news.163.com/special/Watson>.

有三个要素：一是可处理、可达到处理目的的信息集合，二是处理信息的功能系统，三是确定的处理目标。这里，第一要素是一个特定的信息集合，这个信息集合要满足可处理及与要达到的目标相匹配。可处理相对于功能系统而言。不同的功能系统可以处理不同形态、载体或外壳承载的信息，如果不能处理，则需要转换处理，如果不能转换，则不符合条件，即计算的前提不成立。第二要素是功能系统，这个功能系统应与信息集合的形态和数量相一致，与要实现的目标相一致。如果是遗传信息，功能系统是生物体的遗传功能实现系统；如果是认知信息，功能系统是智能生物体的感知、神经和大脑系统；如果是数字化的记录信息或信息系统的自有态信息，则是计算机信息系统，模拟记录信息需要特殊的功能系统。第三要素是目标或目标集合。任何计算必须有确定的目标，否则就是盲目处理活动，要么没有存在的必要，要么就是浪费资源。

信息处理要素构成了必要的信息处理体系架构，图 2.7 是信息处理体系架构的一般抽象。

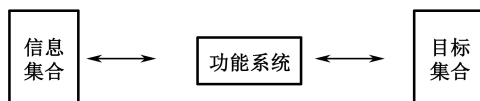


图 2.7 信息处理体系架构

从图 2.7 可以导出信息处理能力（ PF ）一般表达式：

$$PF=(A_0+A_I+A_a)*(F_v+F_t+F_s+F_p+F_r+F_e) \cdot (O_i+O_e) \quad (2.6)$$

其中：

A_0 表示信息集合 A 相对于处理目标的完备性。

A_I 表示信息集合 A 相对于功能系统的可处理性。

A_a 表示信息集合 A 相对于自身语义逻辑关系的结构显性程度，语义逻辑关系结构显性是指信息集合中的语义单元间逻辑关系可理解、可处理的表示。

F_v 表示功能系统 F 相对于信息集合 A 所拥有的信息数量的处理能力。

F_t 表示功能系统 F 相对于信息集合 A 所拥有的信息类型的处理能力。

F_s 表示功能系统 F 相对于信息集合 A 所拥有的语义结构表达方式和深度

的处理能力。

F_p 表示功能系统 F 满足目标集合 O 需求的处理能力。

F_r 表示功能系统 F 根据目标集合 O 的需求得出的处理结果。

F_e 表示功能系统 F 满足目标集合 O 需求提供的处理能力的经济性, 与支付能力的关系或费用的合理性。

O_r 表示目标集合 O 对确定目标表达合理性的能力, 即是否真实反映需求并适应信息集合和功能系统的特征。

O_c 表示目标集合 O 相对于任务提出者 (机构或个人) 需求表达的正确性或全面性的能力。

2.3.7 信息处理能力的演进

式 2.6 显示信息处理能力由 3 个部分 11 个指标构成, 信息处理能力的演进就是这些指标所代表的不同能力发展的过程。对这 11 个指标所代表的能力发展路径简要分析如下。

首先是信息集合的 3 个指标。 A_0 是已有信息集合的完备性指标, 这是最基础的指标, 信息集合不完备, 信息处理的目标不能达成。不同的信息处理目标, 信息集合的完备性程度差异很大。对于转换性处理, 处理的对象就是已有的信息集合, 这样的处理类型, 信息集合必然是完备的。对于其他类型的, 信息集合完备度既是相对的, 又是绝对的。完备度的相对性包括三个方面, 一是由于需求的时间动态性, 需求本身随着认识或事件的发展而发展, 在前一时间点的完备度是 80%, 到后一时间点如果信息集合没有新的增加, 完备度可能只有 50%; 二是由于需求的空间动态性, 相对于目标集合的逻辑完备度, 取决于在同一时间段可能存在的需求集合与表达出来的需求集合是不同的, 满足度显然就不同; 三是相对于既存记录信息的完备度是动态的, 在不同的时间段, 既存记录信息在变化中, 同一问题集合可拥有的信息集合与实有信息集合间的完备性在动态变化。完备度的绝对性同样是这三个方面, 对一个具体的、正在处理的需求, 完备度就是绝对的。信息集合相对于完备度是一个动态变化, 持续演进的过程。演进的阶段、发展的台阶基于信息收

集的能力、支付能力、记录能力和移动能力的信息集合完善能力，更广泛一点，还包括认识、制度约束及其他约束。一般来说，一个信息集合面对的需求是时间的函数。需求面向时间轴求和，信息集合的完备度就表现在数量上，规模成为信息集合完备度的重要指标。这里要特别说明一点，那就是与第3章信息结构的完备度之间的关系。第3章信息结构的完备度，其参照系是相对应的客观存在，而这里的信息集合完备度的参照系是特定信息处理的目标，两者之间存在本质差异。

A_f 是信息集合的可处理性指标。可处理性和易处理性的能力是指信息集合相对于既有功能系统的适应度，适应度越高，能力越强。可处理性主要是指信息集合中信息的类型、结构化程度等。相对于数值型信息处理系统，数值型信息比较易于处理，无论对什么样的功能系统，结构化程度高的比结构化程度低的易于处理。信息类型的易处理性主要依赖功能系统能处理的信息类型和能力，而结构化程度则是一个长期的演进和积累过程，既依赖于结构化技术的发展，也依赖于成本/效益关系。

A_a 是信息集合结构显性程度的指标，与信息集合可处理性密切相关。语义逻辑关系结构显性是信息集合质量最主要的指标之一。与 A_o 类似， A_a 同样要从对历史需求按时间求和角度讨论。如果信息集合孤立地面对一个一个需求，则是 A_f 的范畴。任何信息集合针对一次需求而存在，是不经济的，也和实践不一致。从历史需求求和的角度看，信息集合先验性地将集合中信息单元间的语义逻辑关系用功能系统可处理的方式显性化，以方便一次次具体的需求处理是必然要求，也是各类信息集合的实践发展道路。

其次是功能系统的6个指标。 F_v 是功能系统能有效处理的信息数量的能力。不管是人工系统还是计算机系统，功能系统的首要性能指标是在给定的时间内能处理多大的量。摩尔定律决定了计算机系统处理能力的快速增长，然而到今天为止的发展历史说明，处理能力的增长始终跟不上信息集合数量的增长。当然，这只是一个一般性结论，在不少具体场景，处理能力大于它所面对的信息处理要求。正是这种总体不能满足处理需求，局部大于处理需求的现象，促使计算模式不断演进，以充分利用计算资源、更好地满足需求。

F_t 是功能系统处理不同类型信息能力的指标。记录信息最早的形态是图形,紧接着是文字和数字,数字的比重在记录信息的总量中始终很小。但是,算盘、机械、真空管、半导体、集成电路所代表的信息处理能力提升,首先在数值型信息处理取得突破,而文字、图像、音视频信息处理能力增长一直滞后于需求。经历了数值型计算能力数千年的缓慢增长和最近 60 多年的超高速增长,其他类型信息处理能力在几千年的缓慢增长后,近几年来进入快速增长时期,但主要依靠计算能力的堆积,技术、方法和理论还没有达到集成电路开创的、依摩尔定律增长的发展阶段。

F_s 是功能系统语义结构表达方式的处理能力和深度的指标。本质上,数值型信息的处理也是对数据间语义关系结构化表达方式的处理,数据库技术中的数据字典、元数据等就是其具体实现方式,已经达到了高度结构化和高性能处理的水准。文字、图像、音视频信息的结构化、结构表示的显性化自身的理论方法不成熟,功能系统同样受制于此,这两者是相辅相成的。

F_p 是功能系统能否及时满足需求的指标。这是对功能系统处理能力的综合评价。任何需求都有一定的时间要求,超出了这个时间点,成功地处理也没有现实的价值。但是,用户往往会根据先验的评价,确定需求的提出与否及时间的要求,这也是现实需求和潜在需求差距的一个重要原因。在认知信息的处理中,同样存在这一情形,只能主体通常采用简化或放弃。

F_r 是功能系统处理结果可用性的指标。这是一个微观指标,根据一个个特定需求结果是否可用作判断。从全局看,可用性是一个关联指标,是否可用,与信息集合本身的质量、需求表述的确切程度高度关联,但最终的评价主要落在功能系统上,因为信息集合和需求表示是给定的。

F_c 是功能系统完成需求成本效益的指标。成本效益指标要求功能系统不仅能够具备能力,还要与支付能力和成本的合理性相适应。这一指标在遗传信息和认知信息处理过程中不存在。

最后是目标表达方面,有 2 个指标。 O_r 和 O_e 是关于目标表达能力的指标,前者指表达的科学性,后者指表达的系统性或全面性。在过去的研究中,

经常忽略了需求本身的不足，而需求表述能力实际上是一项信息处理的起点，起点出现偏差，结果不出偏差是偶然现象。需求提出者往往不是信息和信息处理技术的专业人员，有时候需求的提出也不能兼顾信息集合和功能系统的实际能力，导致了需求表述在正确性和系统性上的偏差，而演进的要求就是不断达到需求表述与信息集合、功能系统能力的一致性。

在本节的讨论中，为节省篇幅，我们没有专门分析遗传信息和认知信息处理的评价和演进。一般而言，遗传信息的处理过程与记录信息的处理过程在体系架构上是一致的，在评价和发展的 11 个指标上基本不相同。遗传信息的处理信息集合是生物体内遗传信息的全集，是基本不变的，是完备的。遗传信息处理的功能系统是生物体遗传功能系统，也是不变的、完备的。遗传信息的目标表达是固定的，完备的。认知信息的处理不仅体系架构与记录信息一致，在评价和发展的 11 个指标上也是高度一致的。主要的区别在于功能系统的赋予性，即能利用就是生物体的神经系统和大脑；信息集合相对稳定，存在根据任务需求而获得新的信息的动机和行为；目标表示语义逻辑与记录信息大体相当，目标表示的物理形态完全不同于记录信息和计算机信息处理系统。

2.4 形式信息的积累

2.4.1 形式信息积累的定义和类型

定义 2.5: 形式信息积累。形式信息积累是指一个信息集合随时间而不断增加的数量。这是一类特殊的信息运动，信息从不同的地点和时间向一个信息集合运动，形成一个与日俱增、完备程度不断提升的信息集合，这里的完备程度是指相对于存在的形式信息。形式信息积累有两种主要形式：一是生物体认知信息的积累，二是记录信息在特定空间的积累。积累也是信息发展过程中的关键环节。生物体信息积累能力决定了当代生物体的智能水平，记录信息积累能力反映了当代社会智能的发展水平。

生物体认知信息的积累是一个自然过程，依赖于生物体的感知能力、学习能力和存储能力。生物体认知信息积累贯穿于生物体一生，以生命诞生为起点，以生命结束为终点。而生物体认知信息积累能力不仅是一个具体个体生命历程的能力，也是该类生物体历史生命演进的结果。

记录信息的积累是一个社会过程，积累的速度和质量，以及延续的时间依赖于获取能力、存储管理能力和服务能力。记录信息的积累体现在一个时间段社会各个积累实体的能力，而个体记录信息集合的积累也是社会能力的函数。

在一个历史的时间点上，生物体认知信息积累与社会记录信息积累之间存在着相互依存、总和基本相等的正相关。相互依存是指生物体认知信息的积累依赖于对社会记录信息的可获得性、可利用性，而记录信息集合的积累又依赖于生物体认知信息的发展，生物体认知信息又是记录信息集合可获得性的基础。总和基本相等是指如果在信息含义的维度上对所有生物体的认知信息和全部社会积累的记录信息求和，则两者是基本相等的。

社会记录信息积累有两种基本形态，一是图书馆类，即将印刷品或数字资料集中在一个特定的物理空间，供特定的读者群体使用。二是信息网络类，数字化的信息存储在网络的存储设备中，共可利用该网络的群体使用，其中的特例是互联网，供所有可以接入到互联网的主体使用，所称主体可以是生物智能体，也可能是信息设备或系统。

2.4.2 形式信息积累的要素和架构

生物体认知信息和社会记录信息积累的一般要素是：可获得性、可管理性和可利用性。其中，可获得性是指信息积累者（包括生物体和社会记录信息积累主体）信息获取能力，既有自身的获取能力，也有外界对可获得性的约束。可管理性是指积累者对已经获取的信息的管理能力，包括存储能力、组织能力。存储能力是指一个积累者能保存多少信息。生物体能存储的信息依赖于大脑的容量和开发程度，生物学研究的成果是，人的大脑大约有数百到上千亿个神经元。而社会记录信息积累者的存储能力是指其模拟信息和数

字信息存储能力的总和。在人类社会的绝大部分时间，只有模拟信息的存储，计算机发明之后，特别是互联网和传感技术的发展，数字信息的比重不断增加。组织能力是指将积累信息组织为便于利用的形式的能力。对于生物体而言，同类生物体不同个体大脑存储的信息的组织管理模式是相同的，能力有差异，个体脑的能力的不同。除了人类之外，还没有其他生物体能够形成记录态信息，所以，本书所述生物体信息积累特指人类个体信息积累。对于不同社会记录信息积累者，可能拥有的能力是类似的，但实际的组织管理模式和能力是各异的。利用能力是指积累的信息供谁用，怎么用。两类信息积累形态的利用能力均与获取能力和组织能力成正相关。

通过上述分析，我们可以得出信息积累的一般体系架构如图 2.8 所示。

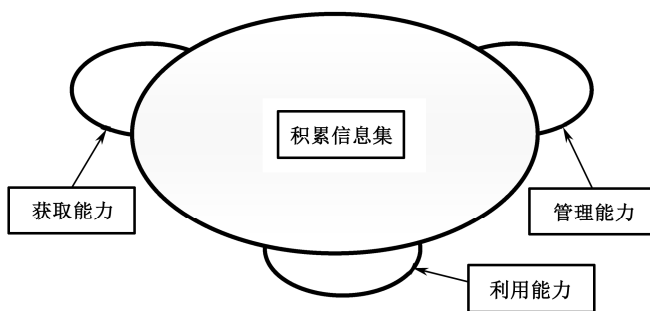


图 2.8 信息积累一般体系架构

任何信息集合信息积累的速度和质量都是获取能力、管理能力和利用能力之函数。但是，一个信息集合实际信息积累的能力还有一系列约束条件。首先是信息域的约束。所谓信息域是指一个社会环境，这个环境中的生物智能体和记录信息使用同样的信息外壳，具有相同的概念体系，换言之，使用相互间无须转换即可理解的信息载体和外壳。一个信息域信息积累的整体水平受制于该信息域的科技文化发展水平。19 世纪 80 年代的英国信息域中的部分信息积累，不管是生物体还是社会记录信息，会有一系列当代物理、生物、天文、机械制造等先进的科学和技术知识，而在世界大部分其他地方则没有

达到这一科学和技术高度的认知或记录信息。其次是地理空间的约束。信息获取或利用在为一个可达的物理空间所约束，尽管在信息获取技术上，人类社会已经取得了巨大的进展，从一个人力可达范畴扩展到物理及信息空间的全球范畴，但信息积累的空间约束依然存在，如果展望地外文明的信息积累，空间约束更是客观存在，如同人类历经万年努力，从自然村走到地球村，结果发现地球村依然是个自然村。信息获取约束的第三个方面是制度、文化、经济等社会约束。政治制度、文化习俗、宗教信仰在不同程度上决定了信息积累的主观意愿和客观来源的不同，直接导致信息积累内容的不同。

2.4.3 形式信息积累的目的和路径

通过形式信息积累，使使用这个信息集合的主体能够得到尽可能完备的信息是人和发展的必然要求，既是积累的动力，也是积累的目的。基于这样的目的，显然任何积累主体必然希望能成为在当时特定范畴中最完备的信息集合。达到这样的目的，发展路径就是不断提升获取能力、管理能力和利用能力。

生物体信息积累的获取能力、管理能力和利用能力是与生俱来的，获取能力和利用能力决定了管理能力，差别体现在前两个能力上。同时代不同个体一生中，实际积累和利用的信息千差万别，原因主要是环境的约束。环境约束有两类，一是生物体生存期的社会认知发展水平，二是不同个体自身在社会的角色所决定的可获得能力、利用能力。可获得能力从主客观两个角度看，首先是主观因素，即经济能力和社会能力。经济能力是指对要求收费的可获得信息的支付能力，社会能力是对已经在同信息域的社会积累信息是否有资格获得。这个资格不仅是地位，在很大程度上是个体在社会中的角色，信息的利益属性决定了谁可以获得具有利益属性的具体信息。

利用能力主要是指生物体认知水平和可用来吸取记录信息的时间。对信息积累整体来说，生物体终身持续不断地积累，这主要指经由五官的感知，并依赖生物体的记忆能力。对于没有记录信息的智能生物体，如猩猩、狗，只有经过感官和记忆的积累。在人类早期，也只有通过感知积累，在文字产

生之后的几千年时间内，超过 95% 的年代中，绝大部分人依然只有通过感知积累。教育普及之后，大部分受教育的人也只有很少的时间通过记录信息积累，互联网普及之后，记录信息在生物体信息积累中的比重才日益增多。实际上对主观因素的分析中也看出了客观因素的重要性。就人类而言，在没有记录信息存在的 300 多万年间，部落相同等级不同个体相比，终其一生，积累的信息差距很少。每代人可能积累的信息量，显然正比于人类文明的进步。当记录信息产生之后，记录信息的生产、流通成本及普及程度，成为不同时代相似社会角色信息积累差距的主要原因。印刷术、机器印刷、平面媒体的发明和发展，不断降低了获得和利用成本，相同社会角色的个体信息积累才得以快速增长。互联网的普及使每一个人积累信息的能力有了极大的增长，很快就会达到记忆力的极限。

生物体认知信息和社会记录信息积累两种形态在互动中发展。而其中生物体认知信息积累对整体信息积累水平提升具有关键的作用。首先是因为社会信息积累的服务对象是以生物体个体信息积累为基础，个体信息积累不仅是某个或若干个个体的水平，更重要的是达到一定水准的个体的数量。其次是社会信息积累的基础是生物体认知信息发展的结果，而发展又是以积累为基础的，没有积累就没有发展。

跨信息域形式信息积累是信息增长的重要模式。从智人形成到语言、文字的发展，人类文明发展历史证明了这一点。

2.5 载体和外壳的演进

信息的神秘就在于含义藏在载体和外壳之后，载体和外壳的演进改变着信息的发展轨迹。我们尚不清楚地球之外文明的发展过程，但在地球文明几十亿年的发展过程中，从自在信息到自有信息、从自有信息到记录信息，其载体和外壳沿着一条清晰的路径演进，而每一次进步都为信息的有效积累和增长开拓了新的路径。每一个具体的载体或外壳在产生的时候，并不知道发展的下一步和终点在哪里，但始终沿着一条恰当的路径走到了今天。全面阐

述这条路径发生发展的规律正是本书的目的，本节只讨论载体、外壳、含义之间的关系及载体、外壳演进的阶段和节点。

2.5.1 载体、外壳、含义的关系及演进模式

载体、外壳、含义是一个具体的客观存在的信息的三要素，自在、自有、记录是信息三种主要存在形态，研究三要素的关系，要分析每一种具体存在形态的实际，要分析相互依存又相互独立的演进路径。

载体和外壳关系的关系具有四种特征。首先是相对性，载体外壳相同与不同的相对性。含义必须具有外壳承载内容，载体作为物理存在的基础，但载体和外壳有时是同一个物体，有时又完全不同。如对一个神经元存储的信息，载体就是外壳，外壳也是载体。如果这个神经元存储的是一个人脸的一个局部的具体特征，而这个局部，如眼睛，需要存储多个具体特征的神经元及其特定的连接来确定，则载体和外壳又不同一了。其次是匹配性，载体和外壳需要物理条件的匹配。假如没有化学递质的匹配，核苷酸之间的结构不能成为遗传信息的外壳，转录和复制就不能实现；而印刷技术将可承载符号的纸张或其他材料结合，才能形成信息发展史上最有代表性的外壳与载体结合的关系。再次是分离性。载体和外壳在一定条件下是可分离的，可分离是信息发展的一个重要条件，若不能分离，信息的增长，特别量的增长（具体见第 4 章）失去了基础。认知信息和记录信息的载体和外壳都有可分离的属性。神经元作为外壳所指代的含义随连接模式与/或刺激强度而改变。可反复读写的半导体芯片载体和外壳始终处于可分离状态。最后是稳定性，是指载体与外壳之间的关系需要保证承载含义的稳定性，这个稳定性既指存在的稳定性，也指承载的含义的稳定性。所谓稳定不是一成不变，而是在功能上满足信息利用、处理、存储、传输的需求。其实，能满足稳定性要求的才能作为含义的载体和外壳，不能满足的就会在发展的历史中淘汰。

外壳与含义的关系也有四个重要特征。首先是可利用性。可利用性是指外壳与含义存在对应关系，相对于利用主体，其含义是可理解、可利用的。遗传基因的外壳可以在转录、复制、翻译等过程中为该生物体所利用，神经

元中表述的含义能为该生物体生存期间利用,记录信息的符号能为理解符号的主体所利用。其次是可表达性。可表达性是指外壳对含义的表达能力。遗传信息的外壳应能充分复杂,表达生命延续和生存过程所有的遗传信息;认知信息外壳应能表达相应生物体认知功能积累的的含义的需求;记录信息的符号体系要能满足记录信息运动全过程所有含义的表述需求。在这里,特别重要的是符号体系要与作为人类认知结晶的概念体系进化相一致。再次是可分离性。最后是稳定性,这里可分离性和稳定性是指外壳与含义既能保持相对稳定性,满足信息运动过程的稳定性需求,还要具备分离的功能,满足信息的增长或功能实现对分离的需求,具体的模式与载体和外壳的相应特征一致。

载体和外壳演进的路径与信息增长和信息空间发展阶段一致,是载体和外壳的演进提供了信息增长和信息空间发展的功能性台阶。分析载体、外壳演进之路,主要的阶梯有:载体——外壳分离,含义、外壳分离,载体可重用,外壳多样化。载体与外壳的分离有四种路径:一是遗传基因信息载体和外壳的分离,使生物体的进化得以实现。二是认知信息载体和外壳分离,使得生物体的智能进化变成现实。三是书面记录信息载体与外壳的分离,实现了人类信息传播的不断增速。四是机器记录信息载体与外壳的分离,这是记录信息处理能力持续增长的一个基础条件。

2.5.2 载体演进和载体与外壳的分离

历经几十亿年,地球的信息载体经过五次飞跃,为信息增长和自我循环的信息空间形成打下了基础。第一次飞跃是自在信息的各类天然物理载体透过生命形成过程诞生了核苷酸作为遗传信息的载体,为信息的演进跨出了第一步。第二次是神经细胞成为认知信息获取、传输、处理和存储的载体,使生物体有可能不断积累信息,通过遗传突变,持续提升生物体获取、处理、利用信息的能力,促进智能的演进。第三次是生物体的信息可以通过肢体语言、特别是语言来表述,是生物体间的交流成为可能,为经验与知识的积累和发展提供了第一个关键的客体。第四次是文字记录的载体形成,从泥板、竹简到纸张,为人类积累的知识大规模跨空间和时间传递创造了条件。第五

次是电子、光子等基本粒子成为依据主体可控的记录态信息载体，为计算机信息处理、远距离快速信息传输和大容量、低成本信息存储奠定了物质基础，更为信息的自在发展提供了物理基础。

遗传基因信息的载体和外壳是可分离的，尽管载体和外壳化学结构相同，但含义由外壳承载。在复制、转录、翻译的过程中，密码子作为含义的承载者，控制着不同的生命进程⁴。

认知信息的载体和外壳是可分离的。神经系统是生物体信息处理器，各种神经细胞是信息的载体，承担感知、传输、处理、汇聚等信息处理功能。认知信息的载体是神经电信号，是载体处理、存储、传输信息不同电位。不同于电子计算机开关逻辑，神经系统的电位是多值逻辑⁵。

当一束方形光线射到兔子的视网膜，落在视网膜的一个特定区域，持续几秒。在这段时间里依次记录 12 种神经节细胞接受到的兴奋性和抑制性信号。每一种细胞都有独特的反应模式，反应的强度也不同，时间和很短，大约只持续十分之一秒神经节细胞的反应不仅滞后，而且只记录入射光下的变化情况。12 种神经节细胞都会输出一个独特的信号，反映视觉神经的一个方面，这个输出信号中不仅有来自双极细胞的兴奋性信号，也夹杂着无长突细胞的抑制性信号，两种信号相互综合、抵消之后才能形成最终的输出信号。

框 2.5 认知信息载体、外壳、含义的连接与分离⁶

语言的载体是震动的空气——声波，外壳是特定个体发出的声音。在一个主体发声、没有噪声的场景中，载体和外壳是同一的，如果不同主体同时发声、噪声与声音重叠，就可能出现同一载体承载多个外壳的情形。语言依

4 分子生物学（原书第五版），[美]Robert F. Weaver 著，郑用琰、马纪、李玉花、罗杰等译，科学出版社，2013。

5 基础生命科学（第二版），吴庆余编著，高等教育出版社，2006 年，第八章和第九章。

6 视网膜上的电影工厂，Frank Weblin and Botond Roska 著，周林文译，大脑与认知，电子工业出版社，2012 年，第 155-161 页。

赖声波，但大部分声波不是语言，各类物体的运动都会产生声波，如风的呼啸、闪电的轰鸣、汽车的行驶等。

载体可重用出现在两种场合：一是认知信息积累过程中，二是计算机等机器信息处理过程中。

从分离和结合时间看，载体、外壳与含义之间又有两种区分度较高的模式：即时性和相对固定模式。遗传基因信息的载体、外壳与含义之间相对稳定，只有在转录、翻译等特定时刻按确定的路径重组，其余时间都保持不变。书面记录信息一旦形成，载体、外壳、含义就固定不变。而认知信息从感知、传输到处理，载体是仓库、外壳随可解释承载的含义变化而变化。

书面记录信息的载体与外壳可分离，这是显而易见的，我们每个人看书，读到了书中的信息，而看的不是纸张，而是文字或插图、表格，甚至可视书本中的声音和图像，按定义，这些是外壳。书面记录信息的载体是纸张其他任何可以承载外壳的材料，从钢铁、丝、棉到石头、泥块，而外壳则是文字、录音、录像、图画、图片等所有可记录人类可理解信息的符号。

机器记录信息的载体与外壳可分离也是显而易见的，如果不能分离，计算机、信息网络、磁盘等设备对信息的计算、传输、存储等功能就无法实现。各类具有计算功能的机器，不管是模拟还是数字，是真空管、半导体还是集成电路，载体是可以区分、表达和存储电信号的器件，而符号则是携带信息的电磁信号。

2.5.3 外壳的演进和外壳与含义的分离

含义与外壳分离是信息增长、信息空间发展的必要条件，是信息三态间转换的基础。外壳与含义的分离除不同态信息转换外还有三种模式：一是基因信息复制、转录等过程中的分离。二是认知信息在形成复合语义、记忆、学习、决策、行为等过程中的分离。三是计算机等机器系统处理过程中的分离。

对信息增长而言，数十亿年外壳的演进有几个关键的台阶。第一步是遗传基因进化为由载体和外壳构成，这种分离，也是遗传基因持续进化、生物

体认知能力不断提升的必要条件。第二步是认知信息载体和外壳的分离和功能日益增强。植物的神经系统中，载体和外壳也是分离的，但结构简单、功能单一。动物的神经系统载体和外壳同步进化，生物电信号的表达能力不断增强，如对颜色的表达和对聚合电信号的综合。第三步是记录信息的载体和外壳分离，形成了承载含义的符号体系。符号体系不仅使外壳的表达能力与认知能力相匹配，更重要的是建立了记录信息与认知信息转换和交互的桥梁，为认知信息和记录信息互动发展创造了条件。

外壳进化的主要标志是表达含义的能力增强。表达能力增强是智能产生和发展的前提，密码子功能的提升、神经细胞对所接受信息含义的表达和聚合能力、语言和文字的产生、机器符号体系的发生和完善都为智能的产生和发展建立了一个又一个关键的台阶。

外壳与含义的分离按时间次序看，首先发生在自在信息向自由信息转换的过程中。植物感知温度，到一定的阈值，触发生长基因，或发芽、或拔节、或开花，外界温度这一客观含义，从植物体感知的光线中分离出来，进入植物体的神经系统。动物经由光线感知一个外部物体，经由视网膜传递到大脑（或其他决定其行为的中枢神经）。光线携带的信息离开自在信息的载体（载体与外壳合一），进入人体后，转入新的载体和外壳，最外层的视觉细胞，而后经过一个长长的过程，一系列的转换，到达大脑相应部位，合成、存储并处理。这些关键的智能行为，都是以外壳与含义的分离为基础的。最后是记录信息外壳与含义的分离，记录信息的产生就意味着这一过程的开始有三种主要形式。一是语言的形成及录音录像技术发展，二是文字的形成及印刷技术的发展，三是模拟信息的记录及模数、数模转换技术的发展。也就是说，在记录信息外壳与含义分离的历史过程中，语言、文字和机器记录是三个最重要的台阶。语言和文字的形成是信息增长历史中重要的里程碑，将在第4章详细讨论。从信息增长的角度看，语言和文字两者，语言的形成是革命性和根本性的，文字是语言形成后信息进化的必然结果。更进一步看，没有语言和文字，机器不能创造，语言文字的机器记录也不可能发生，机器直接记录信息、转换信息也不可能发生。

2.6 本章小结

前面各节分别从转换、移动、处理、积累以及载体、外壳、含义之间的关系等几个方面对信息的运动作了分析，可以看到信息的运动与我们观察到的物质或生命的运动存在重大不同，下面从几个重要的侧面，分析信息运动的特点。

从运动的持续性看信息的运动有两种主要的形态，持续的和间断的；从参与运动的主体看，有三种主要形态，同一主体、同类主体、不同类型主体；从运动的发展阶段看，自在信息是信息运动的起点，信息转换是信息增长的基础，信息移动是信息发展的基础环节，信息处理是信息增长和利用的核心，信息积累是信息运动的阶段性目标，载体和外壳的演进是信息走向高级阶段的必要台阶；从信息运动的目的看，有三类主要目的，生物体自身的生存和发展，社会（或种群，相对于人类社会的其他生物体）的生存和发展，信息的增长。

信息从自在信息通过生物体感知而促使遗传基因的生成和发展；遗传基因的发展促使生物体多样化并推动感知功能和神经系统的进化，生物体认知信息的获取、转换、积累和体系化能力不断增强，最终形成了语言、文字，使个体的知识和信息的积累成为社会的共同财产，加速了信息的积累和增长，拓展了感知自在信息、改造基因信息、提升认知信息的能力，这就是信息运动的目的所在，也是特征所在，规律寓于这样的特征中。

在这一不断发展的信息运动中，存在持续的和间断的具体信息运动。自在信息运动的产生和发送是持续的，与物理世界客观物体的存在和运动相一致。生物体个体基因信息和认知信息的运动是连续的；生物体基因信息的遗传在一般意义上是连续的，除非种群、家族的繁衍中断等特殊情形发生。生物体感知自在信息或记录信息，与生物体生命延续时间一致，以生命的起点

为起点，以生命的终点为终点。如果信息的运动只是以这样的延续和中断方式进行，地球文明只能停滞在“智人”之前的阶段，形不成复杂的语言和文字，更形不成真实反映客观世界的概念体系。从语言、文字和概念体系发展的连续性看，所有个体的认知信息都是不连续的片段。这些不连续的片断经由“社会”这个特殊的存在连接起来。人类历史发展证明，只要承载一个特定文明的社会不覆灭，代表这个文明的语言、文字和由此凝聚的文化就存在。经由社会载体延续的载体、外壳、含义的发展，是信息运动中最具特色的部分，将在第4章详细讨论。

玛雅文明是最古老而充满智慧的一个部落种族，玛雅文明因印第安玛雅人而得名，是美洲印第安玛雅人在与亚、非、欧古代文明隔绝的条件下，独立创造的伟大文明，其遗址主要分布在墨西哥、危地马拉和洪都拉斯等地。玛雅文明诞生于公元前10世纪，分为前古典期、古典期和后古典期三个时期，其中公元3~9世纪为其鼎盛时期。玛雅文明是哥伦布发现美洲大陆之前人类取得的惊人成就。它在科学、农业、文化、艺术等诸多方面，都作出了极为重要的贡献。相比而言，西半球这块大地上诞生的另外两大文明——阿兹台克文明和印加文明，与玛雅文明都不可同日而语。但是，让人们百思不得其解的是，作为世界上唯一一个诞生于热带丛林而不是大河流域的古代文明，玛雅文明与奇迹般地崛起和发展一样，其衰亡和消失。玛雅文明充满神秘色彩。公元8世纪玛雅人放弃了高度发达的文明，大举迁移。创建的每个中心城市都终止新的建筑，城市被完全放弃，繁华的大城市变得荒芜。玛雅文明一夜之间消失于美洲的热带丛林中。

框 2.6 玛雅文明的兴起于消失⁷

在这一不断发展的信息运动中，各不相同的主体参与进来，各自为信息

7 参见百度百科相应词条。

运动的延续和信息的增长作出了独特的贡献。各类物质的存在和运动为信息的运动作出了最初的贡献。所幸存在一类物质，它们的运动与可感知物质运动信息的主体遵循相同的物理规律，如果只存在暗物质、暗能量，那么自有信息和记录信息可能到今天还没有产生。各类生物体为信息的运动作出了最重要的贡献，不仅因为生物体在非生物智能体广泛存在前是自有信息的主要形成者，也不仅因为是通过生物体遗传信息的发展产生并使生物体的智能不断提升，更因为促进了认知信息的发展，而认知信息又进一步带来了信息外壳的进步和记录信息的产生。一个语言、文字、概念体系可共享的信息域是单个生物体信息功能的集成和延伸，使得记录信息的运动超越了个体的局限，通过这样的信息域得到持续性和拓展。社会记录信息的产生和生物智能体的智能的进步，使记录信息、自有信息与自在信息去冗余的信息结构差日益缩小，迈出了信息增长的关键一步（相关定义和讨论将在第3章展开）。具备感知、传输、记录、处理、存储信息能力的机器一诞生，就成为信息移动的又一类主体，提升了生物智能体的信息能力。假如这样的机器系统进一步进化为独立于生物智能体的非生物智能体，将成为自在和自有信息结构的发现者，新的记录信息的创造者，并可能超越人类成为最主要的信息结构完备化的主要承担者。

在这一不断发展的信息运动中，借用不同的力，承担不同的目的，却走向了希望到达的目的地。自在信息的产生和移动是物质世界存在的自然力，不需要附加条件。遗传信息的形成、转录、复制、突变等全过程是生命体自身的生命力。生物体认知信息的感知、传输、处理、合成、使用、存储等运动的全过程也是生命力本身。记录信息运动的全过程动力源自社会力。无论是人把自己的发现将自有信息转换为记录信息，还是各类图书馆积累传播信息的行为；无论是书籍、杂志的出版发行，还是会议报告、数据库建设；无论是目录索引编制，还是维基百科；无论是搜索引擎的发展和利用，还是互联网用户向网络上传自己的信息；不管看起来是个体的行为还是看起来是社会的行为，实际上都是由个人和社会合在一起的社会力量推动信息运动。从信息运动的目的看，有三类主要目的，生物体自身的生存和发展，社会（或种群，相对于人类社会的其他生物体）的生存和

发展，信息的不同类型增长的需求。自然力、生命力、社会力和生命体生存和发展需求、社会发展和生存需求、信息增长需求这三种力和三类需求成为信息运动得以实现的客观基础。

从信息运动的发展阶段看，自在信息是信息运动的起点，信息转换是信息增长的基础，信息移动是信息发展的基础环节，信息处理是信息增长和利用的核心，信息积累是信息运动的阶段性目标，载体和外壳的演进是信息走向高级阶段的必要台阶。在信息运动过程中，一个局部的由载体、外壳，包括其中包含的含义构成的信息集合在毁损之后，并不意味着这一信息集合就一定消失了，只有在没有其他信息载体可替代这一集合中的含义时，这一信息集合的全部或部分信息才会消失，如秦始皇焚书坑儒。而在绝大部分场景下，由于别的记录信息或人的认知信息中存在被毁信息集合的部分或全部信息，所以信息本身没有消失。信息转换增加了新的信息载体或外壳，而没有使原来状态信息载体与/或外壳消失。

第3章

信息结构

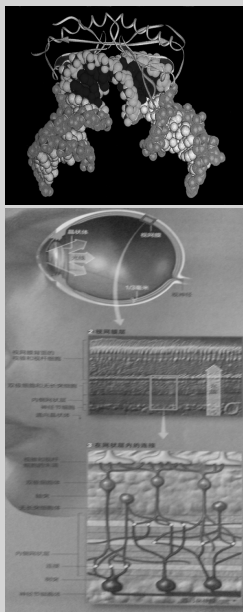
信息结构是信息增长的基石和方向，是理解信息、利用信息的钥匙。

这个奥秘植根于信息的载体、外壳、含义三位一体，按不同的规律、朝着确定的方向跨域运动。

基因如同一架比今天所有计算机还要精密的计算机，基因编码如同一组比今天所有程序还要复杂的程序，控制着每一个生物体发展进程。

神经系统如同一架比今天所有计算机还要精密的计算机，神经元信息处理如同一组比今天所有程序还要复杂的程序，控制着每一个生物体认知进程。

这里的奥秘就是在基因和神经系统中存在关于信息的属性和功能的结构，这种结构为生物体所用。这中结构就是自有信息的隐性结构。



3.1 信息结构的定义

信息是一切客观世界事物的含义，含义通过一定的描述或表示而成为另一类客观存在。对于不同态及不同类的信息，存在多种描述或表示形式，但每一种描述或表示必然显示出信息内在的逻辑结构，因此任何信息集合应该具有对应的信息结构。

基因如同一架比今天所有计算机还要精密的计算机，基因编码如同一组比今天所有程序还要复杂的程序，控制着每一个生物体发展进程。神经系统如同一架比今天所有计算机还要精密的计算机，神经元信息处理如同一组比今天所有程序还要复杂的程序，控制着每一个生物体认知进程。这里的奥秘就是在基因和神经系统中存在关于信息的属性和功能的结构，这种结构为生物体所用。这种结构就是自有信息的隐性结构。每一篇文章的含义不仅由一个个具体的文字、符号表述，还由这些文字和符号的结构表述，这是作者关于这篇文章的隐性结构。

所有的信息集合都实际存在着构成要素间的结构，所不同的是对于利用者而言是隐性的还是显性的。

定义 3.1: 一个信息集合的含义及其内在关系的逻辑体系，称之为信息结构。

定义中的含义包括信息集合中各组件的属性和功能；信息集合是指信息三种存在形态的任何集合，即自有、记录或自在信息的一个空间与/或时间的动态或静态集合；组件是构成一个信息集合的所有可区分的单元。客观存在是指一切发生和发展着的事物，既可以是所有非生命的客体或运动，如一座山、一篇文章、一个图书馆；也可以是有生命的客体或事件，如一棵树、一个人、一个社会事件、一组遗传基因、一个认知过程或认知结果，一个生命体的形成过程等。作为学术研究或艺术创作的成果，这类信息集合的信息结构不仅包括其内容的结构，还包括形成这些内容的过程。关于一个生物体遗传基因的信息结构，不仅是解码该基因组的遗传密码，还包括该遗传基因在

生物体存续期间所发生作用过程的结构性密码。

推论 3.1: 普遍性推论。信息结构覆盖一切客观存在, 具有普遍意义。信息是涵盖物质、能量、生命和信息等自在客体和物质、生命、社会等各类运动(静止是运动的一种方式)的含义, 信息结构也就覆盖了所有这些客观存在的含义和内在逻辑结构的表示或描述。

推论 3.2: 具体性和有限性推论。一个特定的信息结构对应于一个特定的信息集合。信息结构是给定信息集合内容和关系到一种描述或表示(外壳, 也可能是与外壳同一的载体)的映射。信息映射是指客观存在的信息集合到一个信息结构的完备或不完备对应关系。特定信息集合则是关于一个客观存在集合的所有内在含义。一个信息结构是一个特定的信息集合, 根据一定的目的所确定的信息集合要素间的逻辑关系及其体系架构。据此形成的每一个具体信息结构都是有限的, 有限的具体信息结构的总和也是有限的。

从这两个推论可以得出, 任何作为主体理解及利用的信息结构是具体的。任何具体的信息结构都是特殊的, 是一个确定的客观存在的信息含义的映射。非具体的信息结构又是普遍的, 覆盖了所有信息集合, 也就覆盖了所有的客观存在, 从个别到普遍是一个永远没有终点的无限过程。因此具体的信息结构及其总和是有限的, 一般的信息结构是无限的。

推论 3.3: 进化推论。任何具体的信息结构在一定的条件下均具有进化的功能。这是因为: ①功能组件是信息结构的必要构成部分, 具备了进化的基本条件; ②作为信息结构的母体是发展变化的; ③作为其映射产物的信息结构, 只要存在用的需求, 必然进入完善的过程。三者结合, 必然导致信息结构的进化。

3.2 信息结构的类型和形态

如上所定义的信息结构, 依据信息和客观存在的不同, 具有不同的特征和存在形态。本节对主要的类型和存在形态作初步分析, 详细讨论则放在本章以后的相应部分中。

3.2.1 自有信息结构和共有信息结构

信息结构是信息集合的一个基于特定外壳（细胞功能、符号与/或概念体系等）的含义映射。一般而言，相对于利用含义的主体（生物体与/或非生物体），符号或概念体系是专用不是通用，只属于可以理解这一特定符号和概念体系的主体。因此，对于信息结构，从理解、利用的主体特点这个维度划分，存在自有结构和共有结构之分。

定义 3.2: 我们称一个信息结构为自有结构，那么这个信息结构只能为一个主体所利用。

定义 3.3: 我们称一个信息结构为共有结构，那么这个信息结构可以为多个主体所利用。

定义 3.4: 我们称一个信息结构为第一类共有结构，那么这个信息结构可以为同一个信息域的所有主体所利用。

定义 3.5: 我们称一个信息结构为第二类共有结构，那么这个信息结构可以为多个信息域的主体所利用。

定义 3.6: 可以利用、使相同信息结构进化的主体及相关的信息结构一起，组成一个信息域。根据这一定义，一个信息域有两个要素，一是使用相同外壳形成的信息结构；二是使用并使这类信息结构进化的主体。这两个要素，产生两个可能的交叉：一类信息结构与另一类或其他多类信息结构的交叉，一个信息域的主体与另一个或多个其他信息域的交叉。由此得出推论 3.4。

推论 3.4: 信息结构，以及利用、使信息结构进化的主体必然存在交叉。交叉成为信息结构的生长和利用主体能力提升的一个重要，在一定意义上甚至是必要的形式。更加详细的讨论在下一章展开。

每一个具有遗传功能的基因组是典型的自有信息结构，它只能在特定生物体内转录、翻译、复制，而关于遗传基因信息在认知系统中的表述还是在记录信息中的表述，则由于记录信息的共有性，决定了除已经存在于认知信息中，尚没有在记录信息集合中表述的外，均属于共有信息结构。不同阶段、不同内容的认知信息其归属有所不同，没有与外在的记录信息的符号和概念体系对应的，如除记录信息外的感知，属于自有结构；与社会记录信息相一

致的部分，属于第一类共有结构；语言外壳不同，但能够跨信息域交流的则是第二类共有结构。记录信息均属于共有信息结构。自在信息是非表述的信息结构，没有借助任何外壳，也是下面讨论的隐性结构，因此可以看作需要转换的共有信息结构。

3.2.2 隐性结构和显性结构

信息集合所代表的是对应客观存在的所有含义，信息结构是通过一定的转换与/或组织表示出来的部分，而表示既可以是自有信息或记录信息的载体及外壳的功能所致，也可以是专门为理解、利用目的而采取的加工过程所致。因此，无论相对于客体还是主体，信息结构存在隐性与显性之分。

定义 3.7: 隐性结构是指一个具体的信息结构，该结构客观存在，没有经过专门的、用一个或多个信息域的主体共同理解的外壳进行转换或加工，可以为特定的主体为特定的目的所理解和利用，但不为，也不能为此外的主体理解利用或同一主体不同外壳的功能所利用这一定义包含三个要素，一是存在于内含这个信息结构的客观存在中，如一组基因、一段记忆、一本书。二是可理解和可用性，不管是自在态信息还是遗传、认知、记录信息，至少有一个主体的一个目的能理解和利用这个信息结构。三是如果不经以特定外壳为基础的转换加工，不能为同类或同一主体使用。

自在信息、遗传信息、认知信息必然是隐性结构，部分记录信息也是隐性结构。

推论 3.5: 迄今为止，隐性结构是信息结构的主要存在形态。这是显而易见的。无限的自在态信息、众多生物体的隐性结构及结构显性的复杂性和巨大的工作量决定了在显性结构自动化或智能化实现之前，隐性结构是大多数。

定义 3.8: 显性结构是指一个具体的信息结构，该结构经过专门的、用一个或多个信息域的主体共同理解的外壳所进行转换或加工，成为这些主体利用的基础，也成为该信息结构发展的基础。相对于隐性结构，这一定义也包含三个要素，一是经过以特定外壳为基础的并为特定目的转换加工。二是理解这一符号或概念体系的主体都可以利用。三是这个信息结构为智能主体不

断深化把握该信息结构所对应的客观存在的本质或规律创造了提供了持续进展的基础。维基百科、人工智能领域的知识表示是结构显性的一类努力，其结果为信息结构显性做出了贡献，但不是以结构显性为目的，一般而言，也不是一个完整的信息结构。

推论 3.6: 显性结构是共有结构。这是定义规定的。共有结构演进的基础就是信息结构的显性化。

推论 3.7: 显性结构的数量持续增加、质量持续提升。这是不证自明的。近几千年地球文明的发展史在一定意义上是信息结构显性的历史。共有结构是地球文明，也可能是我们所在宇宙文明进步的必要条件。显性结构的增加和完善是信息增长的主要内容。

3.2.3 自在态信息的结构

根据本书第 1 章的定义，自在态信息是客观存在物在环境条件下自身的显现。

定义 3.9: 自在态信息结构是相对于所代表的客观存在的内在含义及其相互间的逻辑关系，也就是客观存在内含的信息结构。第 1 章讨论了信息三态之间的一组关系，但没有讨论信息结构意义上的关系，而信息结构之间的关系是信息三态关系的核心。自在态信息结构是自有结构、隐性结构，本身不具备自我形成其他主体可用的显性结构的可能，不具备自我进化的条件。

推论 3.8: 自在态信息结构是无限的。它也包括自有态和记录态信息结构。从信息集合看，自有态和记录态信息是自在态信息的子集，因为所有的自有或记录态信息首先是自在态，然后才满足自有态和记录态信息的定义要求。同理，自有态、记录态信息的信息结构是自在态信息结构的子集。

推论 3.9: 自在态信息结构是评价自有态、记录态信息结构的最终参照系。判断自有、记录信息结构的正确性、完备性（完备性的定义见 3.6 节），以相应的自在信息结构为基础。信息结构是客观存在含义及其逻辑关系的具体化描述，因此，只有符合该信息结构所表征的客观实在的含义及逻辑关系的才是正确的，只有全面描述了对应客观实在的含义及逻辑关系才是完整的。

定义 3.10:自在态信息集合是指按照一定的逻辑或目的划定的一个范畴,该范畴内的自在态信息构成一个信息集合。这个信息集合可能是概念性的,没有将此集合构成一个实际的存在;也可能是实体性的,如一个数据库、一个文献集合。

3.2.4 自有态信息的结构

定义 3.11:一个自有态信息结构是一个特定自有态信息集合的含义和逻辑关系表征的集合,为特定的主体理解并使用。定义规定了自有态信息结构的边界,对应于特定信息集合;界定了自有态信息结构的的目的性,即为特定的主体所用,特定主体可以是单数,也可以是复数;确定了自有态信息结构的形态是表征而不是描述、表示,该结构具有与所表征信息集合的载体相同的物理、生物属性,是该信息集合内含的含义及逻辑关系。因此,自有态信息结构是与其所表征的信息集合既是一体,又具有独特的功能。独特的功能有两点,一是与自在态信息结构相同的参照系作用,二是在该信息集合转换等运动中具有规定、引导等作用。

推论 3.10:遗传、认知、非生物信息系统三种自有态信息结构的存在形态不同,但在一定条件下具有相同的结构及外壳(符号与/或概念)。例如,作为自有态的遗传信息,其信息结构是有生命的碱基对、碱基对的顺序及决定功能区间的特殊密码。当人类基因组计划实施并破译全部人类基因密码时,作为遗传功能的特定人体的基因信息结构,就和记录信息的文字和概念、存在于记忆能力超强的人脑中以特定的神经细胞结构形态、保存该信息的非生物信息系统中的以0~1模式存在的自有信息结构基本相同。

自有态信息结构既有隐性,也有显性。在三种自有态信息结构中,遗传信息及非生物信息系统中只用于本系统的信息集合的信息结构是隐性、自用的。非生物信息系统中可以为其他信息系统理解并使用的信息结构是显性、共用的。如果一个非生物信息系统具有形成与其他智能体不同表示方式的信息结构,则不属于共有信息结构。如果将这个信息系统的全部功能和构成作

为一个信息集合，则其中关于功能和构成的子集合可能是隐性结构。认知系统中与记录信息相一致并可同为同信息域的主体理解和使用的信息集合的信息结构既有其显性、共用的一面，又有其隐性、自用的一面。说其共性是因为存在共同的符号和概念系统，说其自用是因为表征这些符号和概念的脑细胞结构只能为本人所用。

3.2.5 记录态信息的结构

定义 3.12: 一个记录态信息结构是一个特定记录态信息集合的含义和逻辑关系表征的集合，为特定信息域的主体理解并使用。定义中，特定信息域的数量是不定值，可以是单数，也可以是复数，以记录态信息所使用的符号和概念体系的特征而定。

推论 3.11: 记录态信息集合必然存在显性信息结构，并且是共有的。信息记录的目的一般就是为了具有相同符号及概念系统的信息域的主体共同理解和利用。从信息结构的角度看记录态信息集合，记录态信息集合，不管数量大小，形式不同，必然存在表示其含义及相互逻辑关系的表述。可参阅本章 3.4 节中注解型描述的例子。

推论 3.12: 记录结构的存在形态是以该信息域主体使用的符号与/或概念体系描述或表示的含义及逻辑结构。记录态信息集合的信息结构使用的符号和概念体系与其表示对象并不一定保持相同的符号。例如对大部分图片、图像、音视频、艺术类记录信息的结构表示，不是这些包含这些信息含义的外壳，而是另外的描述语言。本章 3.4 节中的例示从不同角度说明了这一点。

推论 3.13: 记录态信息集合自身展示的信息结构是不完备的。如果需要更加完备的信息结构，要做专门的结构显性工作。记录信息一般只解释其构成含义间的一种或几种逻辑关系，而其内含的信息单元可能的逻辑关系远远大于已经表示的部分。

例如，本书讨论的主概念“信息”一词，可以连接到数以百万计甚至更多的词组或场景，而即使是杰姆斯·葛雷克专门讨论信息的书，《信息，历史、理论、潮流》的目录（该书的显性信息结构）也展示了几十种逻辑关系，全书也只有几千种逻辑关系。

框 3.1 信息一词的逻辑关系

3.3 信息单元

信息结构是有一个个表示一定含义和功能的单元及单元间的连接构成。其实连接也是单元中的构成部分或专门的信息单元，一种功能特殊的信息单元。本节讨论信息单元的定义、属性、构成、类型等一般性主题，详细的讨论在本章后续部分展开。

3.3.1 定义

信息单元是信息结构的基础。理解、构造、进化信息结构，从信息单元开始。

定义 3.13: 信息单元是组成信息结构的基本单位，具有独立含义及功能的语义集合，由相应主体能理解和利用的形式表示，并具备按语义和功能将信息单元组合的一个独立实体。

这个定义包含四个关键要素。首先，信息单元是一个结构化的语义集合。任何信息集合都是语义的，只有结构化的语义集合才是信息结构的单元。其次是具有独立的含义及功能。一个信息单元要有明确的语义及功能，在一个信息结构中具有不可替代的独立作用。再次是这个单元经由主体可理解、可利用，单元间可组合、可分解的方式表示。最后是信息结构中的信息单元是经过特定外壳表示的。表示的外壳可以是活的细胞，也可以是电磁信号，还

可以是符号、概念，只要满足特定主体可识别、可利用，单元可建构条件的都可以作为外壳。

3.3.2 信息单元的一般构成

如此定义的信息单元物理形态和数量不同，但具有如图 3.1 所示的一般构成。

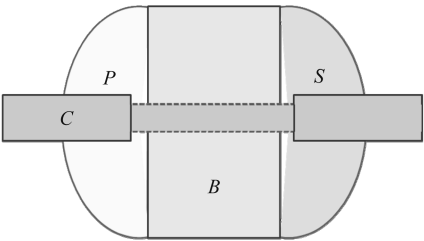


图 3.1 信息单元一般构成

图中，浅绿的 B 是信息单元的基础部分，浅灰的 S 是感知部分，浅红的 C 是连接部分，浅黄的 P 是处理部分。

B 是信息单元的主体和基础，描述一个信息单元之所以能存在的独立、不可或缺的含义、功能。对于信息结构来说，功能也是含义，但在其反映的相应客观存在中，功能与事实存在差别，因此，在信息结构中保留事实和功能的区分，甚至在信息单元中也区分功能为主的单元和事实为主的单元。

S 是信息单元利用和增长的起点。感知并判断来自外部及所属信息结构其他单元与本单元的各种请求，提出相关性等处理要求。感知来自连接，但功能由感知部分实现。例如，连接部分接收到一个与本单元事实表述相关的描述，感知部分决定是否通知处理部分进行判断。

C 是信息单元中最重要的部分，没有连接，等于信息单元不存在。连接部分既包括单元内各部分的连接，还承担本单元与外部及其他单元的连接。连接不仅使单元内组件及本单元与其他单元、甚至跨信息结构的单元建立连

接，更要明确连接的性质，或者说要确定相连双方或各方的关系。因此，连接是信息结构中最复杂、类型最多的一种功能构件。

P 由信息单元中所有处理性功能构成，承担来自 B 、 S 、 C 等部分的处理请求。不同的信息单元处理构件的功能和复杂性各不相同。处理构件主要承担该信息单元利用和增长过程中内部的处理任务，如与相关信息单元的协同、信息单元相关功能和事实调整的判断和执行，等等。

关于信息单元的细化，在第4章和第5章两章中继续讨论。

3.3.3 信息单元的确立

一个信息单元的确立，有两条路径，一是根据信息集中载体的特征，据此确定的信息单元称为物理单元。二是根据信息集中构件的语义特征，据此确定的信息单元称为语义单元。

3.3.3.1 依据载体

定义 3.14: 物理单元是以信息载体的物理特征区分的信息集合。该信息集合的信息结构单元划分的一种形式是与物理单元划分一致。物理单元从大到小可以分成最大物理单元、最小可区分单元和语义连接单元。根据不同的客体和应用需求或处理、存储要求，物理单元可以分解或组合，增加中间层次。

定义 3.15: 最大物理单元是指形式上不能自然分割或自然存在的一个物理客体或由于管理、使用等原因人为切分的一个物理单元。如一篇文章，一本书，一段视频、一段语音，一幅图形或图像。除讨论连接外，一般以最大物理单元为基本信息集合分析其信息结构。

定义 3.16: 最小可区分物理单元是指给予唯一标识符（ID），作为系统中处理和存储的基本信息集合，也可以成为最小或基本信息单元确立的依据。如一篇文章中的一个自然段落，一个网页中的一个自然区隔，一段语音中的一个自然间隔，视频中的一帧，一个立体影像中的一个二维平面，一帧图像、一个图片中的一个特定形状，等等。

定义 3.17: 语义连接单元是指与某一语义标识相关的一组物理单元或语义单元。语义连接单元适用于载体和语义确立的信息集合，可以跨越不同的最大物理集合和语义信息集合，也就能跨越不同的信息结构或信息结构单元。

自在态、自有态、记录态物理单元具有不同的特征。自在态物理单元的意义在于认知主体或认知过程，所以可以隐含在自有或记录态信息的物理单元中。遗传信息最大物理单元显然是一个生物体遗传基因的整体，最小物理单元可以是一个个碱基或一个个蛋白质，而中间层次可以是具有独立功能的一个个基因组或蛋白质组。显然，在遗传基因这类信息态，物理单元和语义单元基本上只重合的。认知信息最大物理单元是一个独立功能的器官或组织。中枢神经系统，大脑，视觉、味觉、听觉、味觉等感觉系统都可以成为确立最大物理单元的依据；最小物理单元是一个个具有特定功能的细胞，而中间态则是组织或器官构成中各类独立物理功能集合。记录态的音频、视频（含图形图片）和文本信息，最大物理单元确立的依据是载体独立的记录信息集合。由于记录的原因及记录的技术手段等条件的不同，独立载体的大小、承载的信息语义类型千差万别。最小物理单元可以是一段音频、一帧视频、一幅图片，也可以是图形、一段文字中每一个可区分音节、物体、文字为一个最小可区分单元。最小物理单元的确立依赖于认知主体或认知需求。对应于社会事件，如一次演唱会、一个案件，自在态信息是事件发生的全过程原貌，自有态信息是相关主体大脑中的记忆，记录态信息则是记录下来的关于该事件的一组音视频及文本。相应的物理单元确立，则采用对应的原则。

3.3.3.2 依据语义

定义 3.18: 语义单元是以信息的语义逻辑关系区分的信息集合。该信息集合的信息结构单元划分的遵循构件间的语义逻辑关系。同理，语义单元也可以分成最大语义单元、最小可区分单元和语义连接单元，也可以根据不同的应用需求分解或组合，形成中间单元。语义单元的确立基于两个需求，一是信息集合的组织；二是应用的需求，这里应用是泛指，从学术研究一直到具体事务处理。因此，语义单元的构件，不管是最大单元、最小单元、还是中间各层次，均应具体场景的不同而不同。

定义 3.19: 最大语义单元是指在一个特定的场景下最高层次的语义单元。这里的场景可以是问题求解过程, 一个语义主题, 一次学习过程, 等等。因此, 最大语义单元是相对的, 在一个场景中是最大语义单元, 在另一个场景可能是中间语义单元。如在一个专门研究视觉功能的场景, 视神经系统就是最大语义单元; 而在研究感觉系统的场景, 视神经系统就是一个二级语义单元。

定义 3.20: 最小语义单元是指在相应的场景中语义不可再分解的信息单元。如文本最小语义单元是词, 汉字具有独立语义的即是字; 音频最小语义单元是一个音节, 或含义区分的一段声音; 视频最小语义单元一帧中的一个语义独立单元; 图形最小语义单元是图形中的一个语义独立单元。最小语义单元在不同的结构中, 存在包含、嵌套、重叠等关系, 甚至是多重包含、多重嵌套、多层重叠等关系。最小语义单元也是场景依赖的, 一个场景下是最小语义单元, 但在另一个场景就可能是中间层次的语义单元, 如一个人的脸部特征在按相片识别时是最小语义单元, 但在以生物特征(如虹膜)识别是就成为中间单元了。这里的场景是广义概念, 包含所有可能产生一个具体场景的情况, 如主体、语义类、任务、事件等。

语义连接单元也是语义单元的组成部分, 与物理单元中的区分主要因语义关系而连接, 以语义单元标识为连接点。

相同的或相似的自在态、自有态、记录态语义单元具有不同的外壳, 不同外壳表达相同的语义, 这是语义信息单元的一个特征。因语义关系确立的信息集合比因物理载体确立的同一个信息集合有更多的单元和单元间的连接和组合, 如作为一门学科的分子生物学这一语义信息单元和作为一本著作的分子生物学。

3.3.4 信息结构中几类主要的信息单元

尽管信息单元确立有两种路径, 这主要是从信息的存在形态和形与义的处理不同要求而区分的。不管是基于物理载体还是基于语义关联, 其信息结构单元所表述的主要是语义关系, 物理关系则相对简单, 主要是地址及地址

标识，地址间物理关系。根据信息单元所表达的内容的性质，可以分为属性单元和功能单元；根据信息单元所含语义的复杂程度，可以分为最小语义单元和复合单元；根据信息单元形成过程及使用过程是否存在时间特征，可以分为静态单元和动态单元。下面分别讨论除最小语义单元外的上述信息单元类型。

3.3.4.1 属性单元和功能单元

属性单元和功能单元将信息结构中所有单元分成两类，属性为主或功能为主。在同一个单元中，将单元组件分成两类，属性描述组件或功能实现组件。

定义 3.21：属性单元。在一个信息结构中主要承担描述属性的信息单元称为属性单元。在本章 3.4 节的三个例示中，均是属性单元。

推论 3.14：在一个信息单元中，主要承担属性描述的构件是属性组件。

定义 3.22：功能单元。在一个信息结构中主要承担功能性任务的信息单元称为功能单元。在本章 3.4.2 节中，如果主要表述基因转录功能，则该单元就是功能单元。

推论 3.15：在一个信息单元中，主要承担表述功能性任务的构件是功能组件。在本章 3.4 节的例示中，每个信息单元中均有功能组件。

由于信息结构中的许多单元既包含属性信息，也包含功能信息，因此，属性单元与功能单元必然存在重合与交叉，即单元中既有属性组件也有功能组件，区分的标准是哪个为主。而根据本节定义的单元结构，几乎所有的单元都包含属性和功能两类组件。

属性单元或属性组件描述的维度分别为本体（形体和语义）、逻辑位置、空间位置、时间、场景五维。这五个维度可以完整描述单元包括的所有语义或物理事实，可以精准地在语义、物理两个世界定位单元，可以连接到可能发生的任何一个具体应用场景。

功能单元或功能组件的描述分为四组：基本功能层、逻辑功能层、建构功能层、任务功能层。

基本功能层是指任何信息单元及信息结构应该具备的，实现逻辑功能、

建构功能和任务功能需要调用的功能。包括感知、连接、定位、处理四个维度。

感知功能是指一切与本单元或本组件相关的事件能及时感知。这里的事件包括语义的、物理的、场景的三个维度，内部的和外部的两个方面。

连接功能是指信息单元或信息结构内部和外部的各种连接。不仅是建立组件、单元、结构之间的物理和语义连接，更要在其他功能单元或组件的协同下，确定连接的性质、成熟度等。而连接的性质主要是确定连接的双方或各方是什么关系。

功能连接依据功能实现的需求，相对清晰，而确定关系的类型和成熟度是连接功能中最复杂、最普遍的部分。主要的关系类型有六类：一是载体的物理关系，二是语义的物理关系，三是语义的层次关系，四是语义的上下文关系，五是场景语义关系，六是除上述三类关系之外的所有语义关联关系。相对于一般意义的关系的复杂性、多样性，一个具体的信息单元或信息结构，要明确的是具体的关系，把下述具体的关系形式化：遗传，时间相继关系；邻接，空间位置关系；伴生，连带出现关系；等同，相同相似关系；义接，语义相伴关系，包括包含、正相关、负相关、交叉；区划，空间重叠分割关系，包括重叠、局部、全局等关系；场景，场景的相同或相似关系；集接，类型间的关系；跨接，不同类型单元间的关系，如声频、视频、文本间的各种关系，语义单元和物理单元之间的关系。上面是一维语义的状态，即只考虑一种语义场景，而一个信息单元或信息结构的连接关系往往事多维的。如同义概念在不同的语义场景中，可能表示不同的含义，甚至相反的含义，可能居于语义关系中的不同位置，认识并处理好关系的多维性特征是信息结构完备的重要任务。

不论何种连接，在其实现方式上，有的是直接连接，有的是间接连接。直接连接是指单元信息结构间需要连接的内容经两个单元间的连接功能实现，不经由信息结构的总体连接功能或其他单元信息结构的中转，后者就是间接连接。在实际的信息结构中，间接连接比重较大。

逻辑功能层是指各层功能实现过程中必需的逻辑推理功能。包括归纳、演绎、类比、判断、匹配等功能。一般而言，在信息结构中，逻辑推理功能

是一个信息结构的共同功能单元，而不是一个信息单元中的内置功能组件。同时，在一个信息结构中的逻辑推理功能单元也不是一个通用的抽象功能，而是针对这个信息结构已经使用、落实到算法的具体功能。如果这个信息结构在完备过程中需要增加或减少逻辑推理功能，则在完备过程中同步进行。

建构功能层是指任何信息单元或信息结构必须具备的，实现信息单元和信息结构不断反映其表征的客观存在的发展变化，逐步走向完备的内在功能。包括组合、分解、归类、约束、进化、管理等功能。其中组合、分解是一组，承担信息结构内单元的合并或分拆功能，如果是一个有若干个子信息结构组成的集合，还承担子信息结构间的分拆和合并功能；归类是对信息结构内单元的组件或信息单元建立新的类型或调整已有类中构成部分的功能；约束、进化又是一组功能，保证信息单元或信息结构在完备过程中按确定的原则规范化地进行，而不是随意的，这些原则因信息结构的不同而不同，有的主要是科学性的约束，有的则是道德性约束。

任务功能层是指任何信息单元应该具备的，实现该信息单元或信息结构承担的任务所需要的功能，包括学习、决策、行为等。这里需要明确的是信息单元或信息结构承担的任务是什么。主要是两类任务，一是映射到信息结构的客观存在所承担的责任或任务；二是信息单元或信息结构自我完备的任务。这两者之间有区分又有联系，自我完备基于对应客观存在的责任或任务，客观存在承担的责任或任务能否完成好，又依赖于其对应信息结构的完备程度。这里所指的学习、决策、行为等功能不是一般意义的无所不包的功能，而是这个信息单元或信息结构承担任务对应的学习、决策、行为。同逻辑功能一样，这些功能都是具体的，3.4 节的例示中将专门对此说明。

3.3.4.2 组合单元

组合单元是一个没有严格定义的稳定概念，若干个本来可以独立或已经独立的信息单元，按照一定的规则或目的组合成一个信息单元。这样的组合单元可能是暂时的，为某个特定的场景、任务而存在，场景、任务变了，这个组合单元就消失了；也可能是稳定的，任务场景变化后，信息结构判断保留这个单元对完备结构更有利，就成为该信息结构中的组成单元。分解是组

合的逆运动，也包括在组合中。

单元的组合可以沿着语义、载体、甚至外壳的线索展开。如从属、邻接、伴生、语义簇、语义网、场景，文本、音频、视频和图形的载体和外壳形式，均可构成单元组合的脉络。单元组合时需要将各相关单元的属性和功能组件整合起来，整合不是简单的叠加，而是根据新单元在信息结构中的位置和功能，按照单元组成的基本规则重新构建。同理，在一个信息单元内部的组件也可能根据一定的规则或目的进行组合。

在二维或三维的静态单元中，存在大量以最小可区分物理单元为基础的语义静态组合单元。这样的组合单元可以表述一个新事件，也可以表述一个新任务。

3.3.4.3 静态单元和动态单元

分析信息单元和信息结构的动态和静态有两个维度，一是单元或单元组件本身的动态性，二是映射到信息结构的客观存在的动态性。对应第一类，静态单元是指单元或组件的地址或描述是固定的，一经形成，只因信息结构完备的内在原因变化，不随任务、场景而变化；动态单元是指单元或组件的地址或描述是一个函数，函数中的值或内容是随任务、场景而变化。对应第二类，静态单元是指单元或组件的地址或描述不随其反映的客观存在的动态变化而变化，变化只有第一类情况下产生；动态单元是指单元或组件的地址或描述随其反映的客观存在的变化而变。不管是第一类还是第二类，静态和动态的区分不在于是否变化，而在于是因为什么而变化。因此信息结构对所有新获取信息展开的区分和标识形成的单元均属于静态单元。系统在调整、修改等学习过程中形成的单元也是静态单元。

3.4 信息单元及信息结构例示

本节选择几个例子来说明不同类型信息单元和信息结构的构成和特性。

3.4.1 二维静态单元的信息结构例示

记录态信息单元结构，在一般意义上，均可以用二维静态信息单元结构描述，音视频记录信息的三维动态特征，按记录时的帧作为物理单元就变成了二维静态。二维静态信息单元的结构形式如表 3.1 所示。

表 3.1 二维静态信息单元的结构形式

	标识	名称	描述
信息单元	3l	名称	描述
基础部分	3lb	名称	描述
组件 1	3lb1	名称	描述
组件 2	3lb2	名称	描述
.....
组件 n	3lbn	名称	描述
连接部分	3lc	名称	描述
组件 1	3lc1	名称	描述
组件 2	3lc2	名称	描述
.....
组件 t	3lct	名称	描述
感知部分	3ls	名称	描述
组件 1	3ls1	名称	描述
组件 2	3ls2	名称	描述
.....
组件 m	3lsm	名称	描述
处理部分	3lp	名称	描述
组件 1	3lp1	名称	描述
组件 2	3lp2	名称	描述
.....
组件 u	3lpu	名称	描述

这里第一列给出信息单元四个部分及其组件，第二列是标识，第三列是名称，第四列是描述，这是一个最简单的示意图。

图 3.2 是一幅街景视频的截图，典型的二维静态信息单元，本节以此作为此类信息单元结构的例示。图中五个已标号物体分别是逆行的机动三轮车、危险物品货车及其发出的声音、出租车、树、大楼。

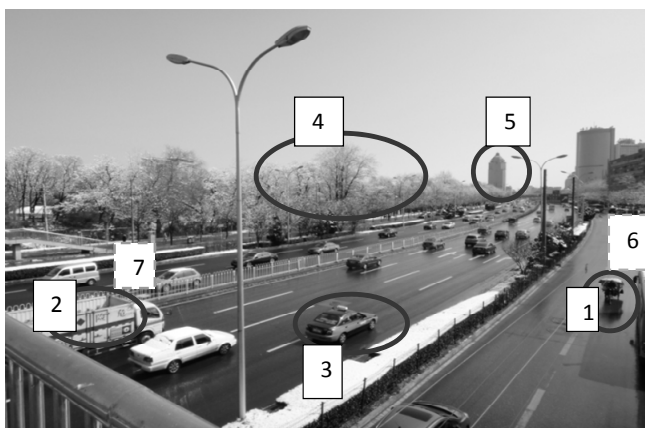


图 3.2 二维静态信息单元例

设 J1 为这一信息单元的标识符，“街景”为其名称，XX 街道××××年××月××日××时××分××秒实景为其描述；设 J1d(f) 为该单元的地址连接函数；设机动三轮车、危险物品货车、出租车、树、大楼、三轮车逆行、危险物品货车声音图中 7 个物品或事件为基本部分的组件，其中三轮车逆行是分析组件，危险物品货车声音为隐含组件，对应的识别标识符为 J1b1, J1b2, J1b3, J1b4, J1b5, J1b6, J1b7；设机动三轮车的连接部分标识为 J1c1，感知部分标识为 J1s1，处理部分标识为 J1p1，空间连接函数为 J1c1c(f)1，时间连接函数为 J1c1c(f)2，语义连接函数为 J1c1c(f)3，空间感知函数为 J1c1s(f)1，时间感知函数为 J1c1s(f)2，语义感知函数为 J1c1s(f)1，处理函数为 J1c1p(f)1，空间处理函数为 J1ply(f)1，时间处理函数为 J1ply(f)2，语义处理函数为 J1ply(f)3；危险物品货车的连接部分标识为 J1c2，感知部分标识为 J1s2，处理部分标识为 J1p2，空间连接函数为 J1c2c(f)1，时间连接函数为 J1c2c(f)2，语义连接函数为 J1c2c(f)3，空间感知函数为 J1c2s(f)1，时间感知函数为 J1cs(f)2，语义感知函数为 J1c2s(f)1，处理函数为 J1c2p(f)1，空间处理函数为

J1p2y(f)1, 时间处理函数为 J1p2y(f)2, 语义处理函数为 J1p2y(f)3; 树、大楼、三轮车逆行、危险物品货车的相应标识和属性类推。其中 $x(f)$ 是一个指代性表述, 代表一种可描述的行为。如 J1c1c(f)1 在不同的场景中可表示为机动三轮车在图中的空间坐标、与图中行人或机动车的距离等; J1c1c(f)2 在不同的场景可表示为图中三轮车的时间及与前后帧的时间差, 或与其他帧时间表示的连接函数, 等等; J1c1c(f)3 在不同场景下可表示为三轮车的语义逻辑族, 如结构、功能、类似产品、产品型号规格, 也可以是本图中的语义逻辑关系, 如速度上限、行驶方向、图中其他物体的语义关联, 等等。据此, 我们可以构成表 3.2, 即街景信息单元的例示性结构。

表 3.2 街景信息单元的例示性结构

	标识	名称	描述
静态街景	J1	××街道××××年××月 ××日××时××分××秒 实景	如图 3.2, 综合功能: 空间关系、时间关系、处理关系、命名规则、地址规则、组合分解规则、……
基础部分	J1b		摘取 7 个物体作为信息单元对象
组件 1	J1b1	机动三轮车	见图 3.2 之 1, 右侧车道逆行
组件 2	J1b2	危险物品货车	见图 3.2 之 2
组件 3	J1b3	出租车	见图 3.2 之 3
组件 4	J1b4	树	见图 3.2 之 4
组件 5	J1b5	大楼	见图 3.2 之 5
组件 6	J1b6	三轮车逆行	见图 3.2 之 6
组件 7	J1b7	危险物品货车声音	见图 3.2 之 7, 视频带声音
连接部分	J1c	某街景既定组件的连接	几个组件均在同一截屏图片中
组件 1	J1c1	J1c1c(f)1、J1c1c(f)2 J1c1c(f)3	机动三轮车空间、时间、语义、事件相关的连接
组件 2	J1c2	J1c2c(f)1、J1c2c(f)2 J1c2c(f)3	危险物品货车空间、时间、语义、事件相关的连接
组件 3	J1c3	J1c3c(f)1、J1c3c(f)2 J1c3c(f)3	出租车空间、时间、语义、事件相关的连接
组件 4	J1c4	J1c4c(f)1、J1c4c(f)2 J1c4c(f)3	圈内树空间、时间、语义、事件相关的连接

续表

	标识	名称	描述
组件 5	J1c5	J1c5c(f)1、J1c5c(f)2 J1c5c(f)3	圈内大楼空间、时间、语义、事件相关的连接
组件 6	J1c6	J1c6c(f)1、J1c6c(f)2 J1c6c(f)3	机动三轮车逆行空间、时间、语义、事件相关的连接
组件 7	J1c7	J1c7c(f)1、J1c7c(f)2 J1c7c(f)3	危险物品货车声音空间、时间、语义、事件相关的连接
感知部分	J1s	某街景既定组件的感知	与实际场景一致的感知
组件 1	J1s1	J1s1s(f)1、J1s1s(f)2 J1s1s(f)3	机动三轮车空间、时间、语义、事件相关的感知
组件 2	J1s2	J1s2s(f)1、J1s2s(f)2 J1s2s(f)3	危险物品货车空间、时间、语义、事件相关的感知
组件 3	J1s3	J1s3s(f)1、J1s3s(f)2 J1s3s(f)3	出租车空间、时间、语义、事件相关的感知
组件 4	J1s4	J1s4s(f)1、J1s4s(f)2 J1s4s(f)3	圈内树空间、时间、语义、事件相关的感知
组件 5	J1s5	J1s5s(f)1、J1s5s(f)2 J1s5s(f)3	圈内大楼空间、时间、语义、事件相关的感知
组件 6	J1s6	J1s6s(f)1、J1s6s(f)2 J1s6s(f)3	机动三轮车逆行空间、时间、语义、事件相关的感知，特别是是否发生交通事故
组件 7	J1s7	J1s7s(f)1、J1s7s(f)2 J1s7s(f)3	危险物品货车声音空间、时间、语义、事件相关的感知，特别是是否存在安全隐患
处理部分	J1p	某街景既定组件的处理信息	连接、感知、判断、行为相关内部处理，不包括交外部的处理
组件 1	J1p1	J1p1p(f)1、J1p1p(f)2 J1p1p(f)3	机动三轮车空间、时间、语义、事件相关的处理，特别是速度和周边人、车、物的距离
组件 2	J1p2	J1p2p(f)1、J1p2p(f)2 J1p2p(f)3	危险物品货车空间、时间、语义、事件相关的处理，特别是车况和速度
组件 3	J1p3	J1p3p(f)1、J1p3p(f)2 J1p3p(f)3	出租车空间、时间、语义、事件相关的处理

续表

	标识	名称	描述
组件 4	J1p4	J1p4p(f)1、J1p4p(f)2 J1p4p(f)3	圈内树空间、时间、语义、事件相关的处理
组件 5	J1p5	J1p5p(f)1、J1p5p(f)2 J1p5p(f)3	圈内大楼空间、时间、语义、事件相关的处理
组件 6	J1p6	J1p6p(f)1、J1p6p(f)2 J1p6p(f)3	机动三轮车逆行空间、时间、语义、事件相关的处理，同组件一
组件 7	J1p7	J1p7p(f)1、J1p7p(f)2 J1p7p(f)3	危险物品货车声音、空间、时间、语义、事件相关的处理

上述单元内物体区分和表的描述只具示例作用，服务于特定目的的该信息单元结构和描述应侧重于特定场景及其延伸和分析，并根据需要确定详尽度和颗粒度。从详尽度看，图中车这一大类中还有轿车、货车、出租车、轿车；树和房子都是复数，可以进一步扩展；除了列举的 7 个组件外，图中还有很多可区分的物体、事件和隐含组件，详尽度需要根据任务或语义清晰度的要求来确定。从颗粒度看，现在的描述只是例示性分析一个道路状态的层次，离最小组件-信息点还有一定差距，把表中组件描述的一个内容作为一个组件，将其中的内容作为信息点细化，才达到足够的颗粒度。如组件 7，可以把声音特征值、空间相对位置、速度、车的型号和性能、营运许可和要求、运输的物品、运输的路线、当前位置与预期位置的差距、运营各方的反馈功能等作为信息点描述，就可以实现为危险货物运输监控提供必要信息的功能，也可以成为假设的该危险货物运输显性信息结构中的一个构成信息单元。

3.4.2 三维静态跨信息态信息结构的示例

除记录信息外，绝大部分客观事物的静态存在都是三维的，这里选取神经系统视觉功能作为这类信息结构的例示。

从信息结构看，神经系统视觉功能已经是一个十分复杂功能的显性，由

成千上万的信息单元构成。如果以构成的基本单元为例示，可能涉及过于专业的内容，既不利于绝大部分读者的理解，也不利于系统性地解释这类信息结构显性和完备需要面对的问题。表 3.3 列举神经系统视觉功能信息结构 1~3 层不完全的组件，4 层以下的不包含在本表中。

表 3.3 神经系统视觉功能信息结构组件（部分）

一级类	二级类	三级类	标识	描述
视网膜			R1	总体部分，略
	色素上皮层		R11	分别从 B、S、C、P 四个部分进行格式化描述，下同
		色素上皮细胞	R111	*
	光感受器细胞		R12	
		视杆细胞	R121	*
		视锥细胞	R122	*
		视色素	R123	*
	视网膜神经回路		R13	
		双极细胞	R131	
		神经节细胞	R132	
		水平细胞	R133	
		无长突细胞	R134	
		水平连接	R135	*
		垂直连接	R136	*
感光换能			R2	光感受器细胞的感光换能
	视杆细胞感光换能机制		R21	
		视紫红质	R211	视蛋白、视黄醛
		视紫红质光学反应	R212	*分解、合成
	视杆细胞电位		R22	
		静息电位	R221	*
		感受器电位	R222	*
		部分略

续表

一级类	二级类	三级类	标识	描述
信息处理			R3	视网膜信息处理
	视网膜神经细胞 光反应特征			
		双极细胞	R311	
		部分略
视觉分析			R4	中枢神经的视觉分析，总体 部分，略
外侧膝状体			R5	
	大细胞层		R51	
	粒状细胞层		R52	
	M 型节细胞		R53	*
	P 型节细胞		R54	*
	非 P 非 M 型节 细胞		R55	*
	LGN 神经元		R56	
		膝状体神经元	R561	
		中间神经元	R562	
初级视皮层			R6	
	光反应特征		R61	
		简单细胞	R611	
		复杂细胞	R612	
	朝向柱		R62	
	双眼细胞		R63	*
	超柱		R64	
	平行信息处理		R65	
		大细胞通路	R651	*
		小细胞通路	R652	*
		粒状细胞通路	R653	*
纹状外视皮层			R7	
	背侧通路		R71	
	腹侧通路		R72	
		部分略

本表内容主要基于于龙川主编，北京大学出版社出版的《神经生物学》，并参照了环球科学杂志社编，电子工业出版社出版的《大脑与认知》等书籍和文章。

从表 3.3 中选取带 “*” 的组件说明这类信息单元结构的形成模式。

下面对表 3.4 中基础部分的描述函数分别做一个简要的示例性分解。

表 3.4 色素上皮细胞的结构描述

名称	标识	描述类	描述函数
色素上皮层	R11	基础部分(B)	文字描述：结构 R11bd1(f)，功能 R11bd2(f)； 图形描述：R11bv1(f)； 图形注解：R11bn1(f)
		感知部分(S)	文字描述：结构 R11sd1(f)，功能 R11sd2(f)； 图形描述：R11sv1(f)； 图形注解：R11sn1(f)
		连接部分(C)	文字描述：结构 R11cd1(f)，功能 R11bc2(f)； 图形描述：R11cv1(f)； 图形注解：R11cn1(f)
		处理部分(p)	文字描述：结构 R11pd1(f)，功能 R11pd2(f)； 图形描述：R11pv1(f)； 图形注解：R11pn1(f)

其中物理地址表示该表在一个存储系统中的物理地址，成熟度表示该内容的可信度（见表 3.5～表 3.8，图 3.3～图 3.4）。

表 3.5 色素上皮细胞基础部分结构描述

标识	物理地址	成熟度	描述体
R11bd1(f)	A0010101	0.9	防止光线的反射，消除来自巩膜的散射光线，强光照射视网膜时伸出伪足样突起，避免光感受器受到过度光刺激；为光感受器细胞提供来自脉络膜的营养，吞噬光感受器细胞外段脱落的膜盘和代谢产物；含有丰富的维生素 A，维持光感受器细胞视色素的正常代谢

表 3.6 色素上皮细胞基础部分功能描述

标识	物理地址	成熟度	描述体
R11bd2(f)	A0010102	0.9	色素上皮细胞是色素上皮层的主要构成部分，色素上皮层不属于神经组织，但是视觉功能的重要组成部分，位于视网膜的最外侧，内侧与光感受器细胞连接

表 3.7 色素上皮细胞基础部分图形描述

标识	物理地址	成熟度	描述体
R11bv1(f)	A0010103	0.9	见图 3.3，物理地址：A0010201，A0010202，A0010203

表 3.8 色素上皮细胞基础部分图形注解

标识	物理地址	成熟度	描述体
R11bn1(f)	A0010104	0.8	见图 3.4，物理地址：A0010204，A0010205，其中 a 是绘图式注解图，b 是影像式注解图

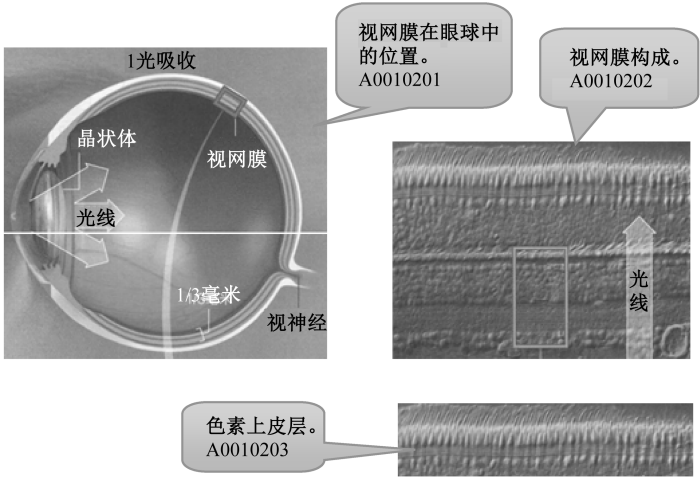


图 3.3 色素上皮层及其在视网膜的位置，视网膜在眼球中的位置

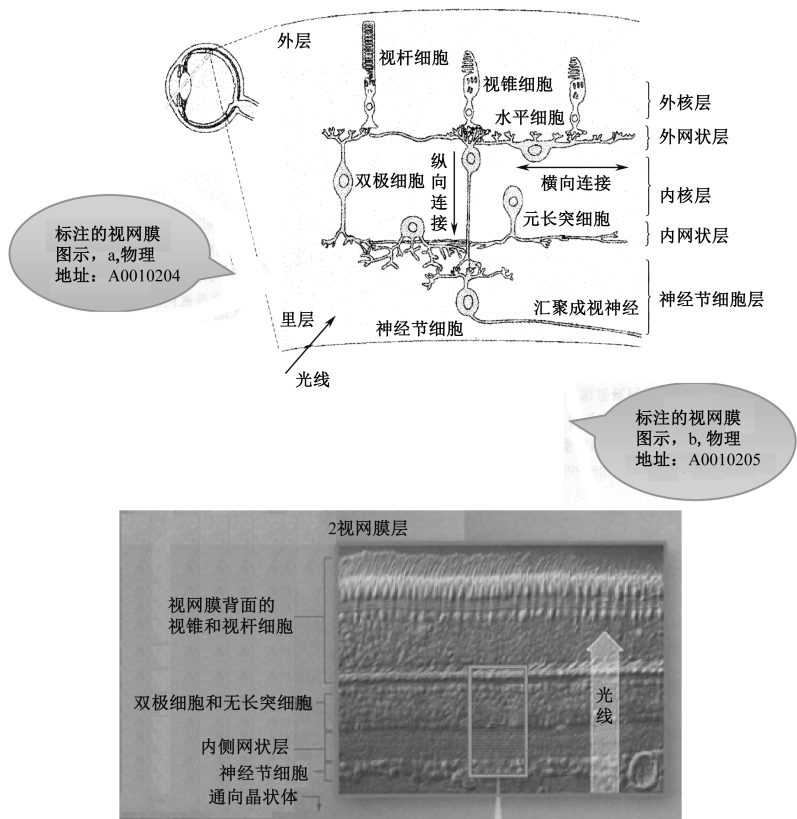


图 3.4 色素上皮层及其在视网膜的位置，视网膜在眼球中的位置

下面对图 3.4 中视杆细胞感知部分描述做例示性展开（见表 3.9～3.13，图 3.5～3.6）。

表 3.9 视杆细胞的结构描述

名称	标识	描述类	描述函数
视杆细胞	R121	基础部分 (B)	文字描述: 结构 R121bd1(f), 功能 R121bd2(f); 图形描述: R121bv1(f); 图形注解: R121bn1(f)
		感知部分(S)	文字描述: 结构 R121sd1(f), 功能 R121sd2(f); 图形描述: R121sv1(f); 图形注解: R121sn1(f)

续表

名称	标识	描述类	描述函数
		连接部分 (C)	文字描述: 结构 R121cd1(f), 功能 R121bc2(f); 图形描述: R121cv1(f); 图形注解: R121cn1(f)
		处理部分(p)	文字描述: 结构 R121pd1(f), 功能 R121pd2(f); 图形描述: R121pv1(f); 图形注解: R121pn1(f)

表 3.10 视杆细胞感知部分结构描述

标识	物理地址	成熟度	描述体
R121sd1(f)	A0020201	0.9	视杆细胞可分为三部分, 由外向内依次为外段、内段和终足(突触终末)。外段呈圆柱状, 有一些重叠、排列整齐的圆盘状结构, 称为膜盘, 是视色素集中的部位, 视色素是产生视觉的物质基础。视杆细胞只有一种视色素, 称为视紫红质。每个视杆细胞外段中重叠有近千个膜盘, 每个膜盘中约含 100 万个视紫红质分子

表 3.11 视杆细胞感知部分功能描述

标识	物理地址	成熟度	描述体
R121sd2(f)	A0020202	0.9	进入视网膜的光量子碰到视紫红质分子, 视黄醒发生构象变化, 转变为全反型视黄醒, 并与视蛋白分离, 同时视蛋白激活, 通过与其偶联的 G 蛋白, 激活下游效应酶, 诱发视杆细胞产生感受器电位, 一个光量子被视紫红质吸收后即可使 11-顺视黄醒变为全反型视黄醒, 在暗处视物时, 视紫红质既有分解, 也有合成, 这是在暗处能视杆细胞的静息电位和感受器电位, 光照时, 外段膜膜盘的视紫红质发生光化学反应, 引起膜盘上的转导蛋白活化, 激活磷酸二酯酶, 将细胞外段胞浆中的 cGMP 分解成 5'-GMP, cGMP 浓度降低, 致使外段膜上 cGMP 门控 Na ⁺ 通道关闭, Na ⁺ 内流减少, 胞膜超极化(图 12-10), 因此视杆细胞具有超极化感受器电位。感受器电位电紧张性地扩散到细胞终足, 引起谷氨酸释放量发生改变, 其下游的双极细胞产生超极化或去极化电位改变, 进一步引起神经节细胞放电频率发生变化, 逐级传递至视皮层产生视觉

表 3.12 视杆细胞感知部分图形描述

标识	物理地址	成熟度	描述体
R121svl(f)	A0020203	0.9	见图 3.5，物理地址：A0020501

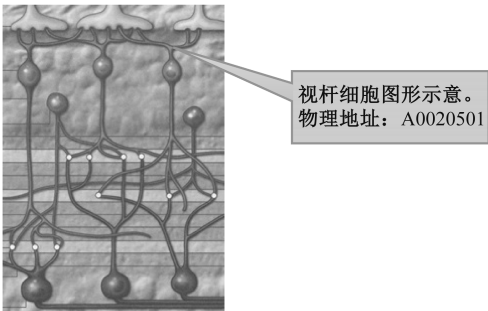


图 3.5 视杆细胞图形描述示例

表 3.13 视杆细胞感知部分图形注解

标识	物理地址	成熟度	描述体
R121sn1(f)	A0020104	0.9	见图 3.5，物理地址：A0020502，A0020503

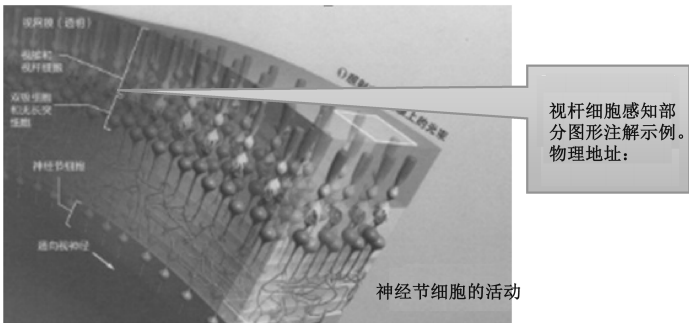


图 3.6 视杆细胞感知部分图形注解示例

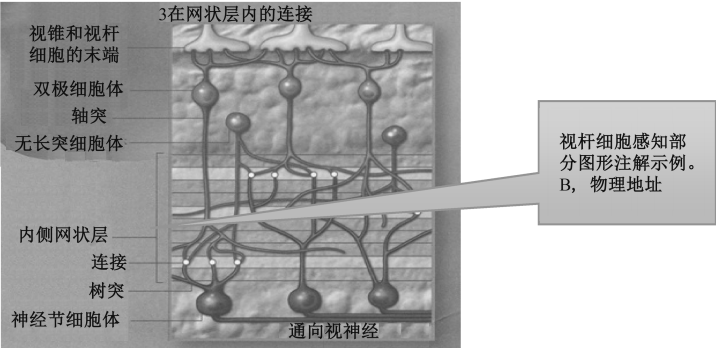


图 3.6 视杆细胞感知部分图形注解示例（续）

下面对表 3.14 中双极细胞连接部分描述做例示性展开(见表 3.14~3.17, 图 3.7)。

表 3.14 双极细胞的结构描述

名称	标识	描述类	描述函数
双极细胞	R311	基础部分 (B)	文字描述: 结构 R311bd1(f), 功能 R311bd2(f); 图形描述: R311bv1(f); 图形注解: R311bn1(f).
		感知部分(S)	文字描述: 结构 R311sd1(f), 功能 R311sd2(f); 图形描述: R311sv1(f); 图形注解: R311sn1(f).
		连接部分 (C)	文字描述: 结构 R311cd1(f), 功能 R311bc2(f); 图形描述: R311cv1(f); 图形注解: R311cn1(f).
		处理部分(p)	文字描述: 结构 R311pd1(f), 功能 R311pd2(f); 图形描述: R311pv1(f); 图形注解: R311pn1(f).

表 3.15 双极细胞连接部分结构描述

标识	物理地址	连接类型	连接类型	连接类型	连接成熟度	描述体
R311sd1(f)	A0030301	A0030301	A0030301	A0030301	0.9	光感受器电位产生后，通过影响光感受器细胞的递质释放，引起下游的双极细胞发生超极化或去极化等级电位，水平细胞发生超极化的等级电位，产生动作电位，这些动作电位作为视网膜的输出信号进一步向中枢传递

表 3.16 双极细胞连接部分功能描述

标识	物理地址	成熟度	描述体
R121sd2(f)	A0020202	0.9	进入视网膜的光量子碰到视紫红质分子，视黄醛发生构象变化，转变为全反型视黄醛，并与视蛋白分离，同时视蛋白激活，通过与其偶联的 G 蛋白，激活下游效应酶，诱发视杆细胞产生感受器电位，一个光量子被视紫红质吸收后即可使 11-顺视黄醛变为全反型视黄醛，在暗处视物时，视紫红质既有分解，也有合成，这是在暗处能视杆细胞的静息电位和感受器电位，光照时，外段膜盘盘的视紫红质发生光化学反应，引起膜盘上的转导蛋白活化，激活磷酸二酯酶，将细胞外段胞浆中的 cGMP 分解成 5'-GMP，cGMP 浓度降低，致使外段膜上 cGMP 门控 Na ⁺ 通道关闭，Na ⁺ 内流减少，胞膜超极化（图 12-10），因此视杆细胞具有超极化感受器电位。感受器电位电紧张性地扩散到细胞终足，引起谷氨酸释放量发生改变，其下游的双极细胞产生超极化或去极化电位改变，进一步引起神经节细胞放电频率发生变化，逐级传递至视皮层产生视觉

表 3.17 双极细胞连接部分图形描述

标识	物理地址	成熟度	描述体
R121sv1(f)	A0020203	0.9	见图 3.7，物理地址：A0020501

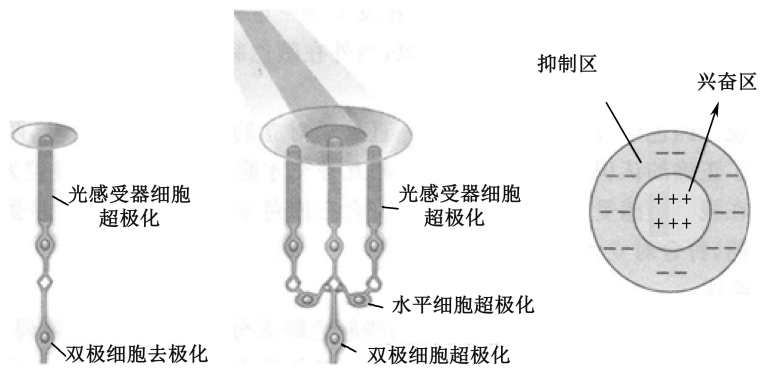


图 3.7 双极细胞图形描述示例

3.4.3 三维动态跨信息态信息单元的示例

启动子及转录起始过程的信息，按迄今为止的理解，应该有四种存在形态：一是自在态信息，脱离主体的遗传基因，如实验室中从某生物体中提取的一个基因片段；二是自有态遗传信息，如某活着的生物体内的遗传基因中的启动子，正在为转录而发挥作用；三是学习研究遗传基因的人的大脑中关于启动子的信息，图像、文字说明等；四是记录态信息，存在于各种文献中关于启动子的描述。

这里将自有态和记录态信息关于启动子及转录起始的信息单元综合在一起讨论其一般构成。如图 3.8 所示，启动子是基因中的一段，作用是确定转录的起点并与聚合酶一起为转录提供初始态。

从自有态基因活动的角度看，图 3.9 的四个部分构成了启动子及转录起始这一信息单元。转录以起始、延伸和终止三个步骤发生。起始包括 RNA 聚合酶结合到启动子上、局部解链及形成最初的磷酸二酯键。在延伸过程中，RNA 聚合酶沿着 5'——3'方向将核糖核苷酸连接在一起延伸 RNA。最后，在终止时，聚合酶和 RNA 产物脱离 DNA 模板。

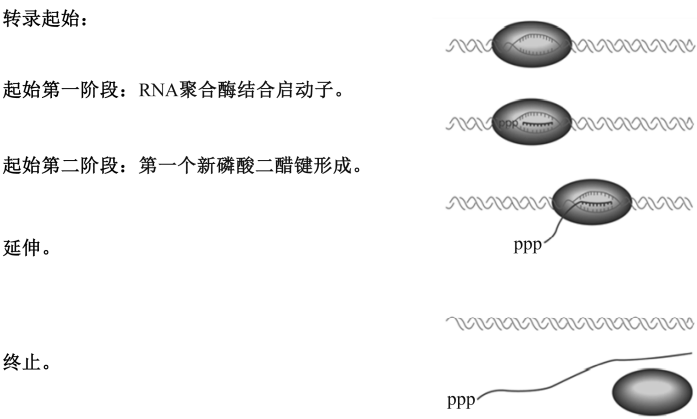


图 3.8 启动子的位置及作用示意图

(来源: 《分子生物学 (原书第五版) 》, 第 37 页)

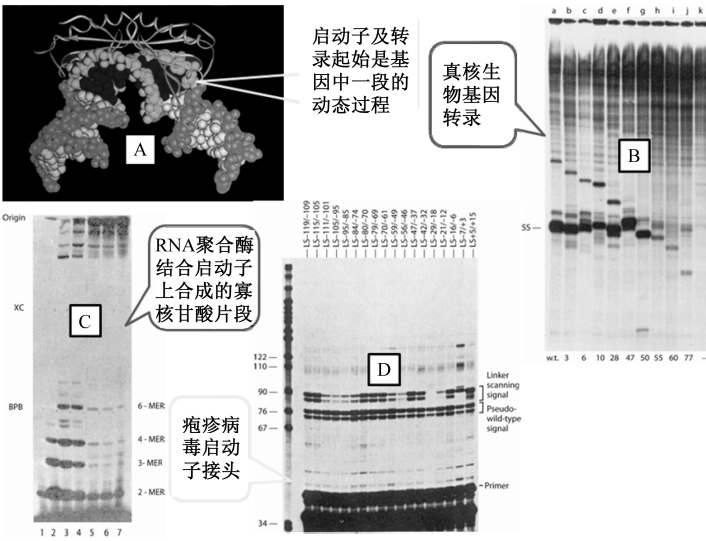


图 3.9 自有态信息启动子和转录起始信息单元的图示

(来源: 根据《分子生物学 (原书第五版) 》, 第 122、259、262、273

页相关内容组合而成)

图 3.9 有四个部分，图片 A 是 TBP-TATA 框复合物的 X 射线晶体结构。TATA 框在大多数 RNA 聚合酶 II 所识别的真核生物启动子中位于转录起始点上游 25-30bp 处。TBP 是 TATA 框结合蛋白，分别存在于 I、II、III 类前起始复合物的 SL1、TF II D 及 TF III B 因子中的一个亚基，与含有 TATA 框的 II 类启动子的 TATA 结合。图片 B 是 III 类启动子一个具体转录过程的记录，说明启动子的转录结果的影响。图片 C 是 RNA 聚合酶结合启动子上合成的寡核苷酸片段。图片 D 是疱疹病毒 tk 启动子的接头扫描突变效应。

下面我们把图 3.9 的内容与信息单元的一般构成进行对照。假设将该信息单元标识为 G1，则图片 A、B、C、D 中影像部分构成 G1 的浅绿部分，描述启动子及转录起始过程的事实、功能和运动。A 包含了给定生物体与启动子和转录起始过程全部相关的事实信息和功能信息，启动子的构成、与其他基因和基因活动过程的关系、维持生命的功能、与聚合酶融合的功能、转录过程中退出的功能，等等。B、C、D 则分别记录了基因转录起始过程的一个瞬间。反映了转录过程中不同泳道电泳信号的强弱、连续性或结果等状态（事实）。

在 G1 所代表的信息单元中，包含了必要的处理功能。例如，保持启动子在生命过程中发生各种突发状态而不变，保证在转录发生时与聚合酶的融合，形成闭合和开放启动子重合体，与聚合酶结合形成最初的几个核苷酸，启动子清除出转录过程等都是处理功能的范畴。这也很好地说明了动态的信息集合一定包含处理功能，相应，动态的信息单元，也应包含处理功能。

在 G1 所代表的信息单元中，同样包含了必要的感知功能。启动子维持生命活性及保持功能的过程中，必须具备相应的感知功能，以正确辨别需要摄取的物质，应对基因活动过程中的各种事件。在转录起始过程中，要正确感知来自聚合酶的信息，要正确感知转录物的长度及稳定性，为处理功能提供信息。

在 G1 所代表的信息单元中，连接是其不可或缺的功能。在图片 A 中，存在着关于启动子生存和正常发挥功能的全部连接关系，在图片 B、C、D 中存在着与转录起始过程相关的全部连接关系。

实际存在的启动子及转录起始过程中，处理、感知、连接的功能经常由

相同的细胞完成，但对于信息结构基本组成部分的信息单元，区分这种功能是很重要的，这种区分将同一客观存在的不同态信息结构之间建立了可比较的通道。

实际上，图 3.9 也可以看作一个关于启动子和转录起始过程的记录态信息单元描述的图像部分之一。作为记录态信息结构的信息单元，显然这是一个不完备信息单元。它只是一个生物体启动子及三个转录起始过程的实体影像，而不是抽象的概念意义上的启动子及转录起始过程。对于一个生物体的启动子及该生物体的一次转录过程来说，三个图片只是三个瞬间，没有全过程系统描述。因此，关于启动子和转录起始过程的记录态信息单元，应该具有如下的结构：

单元标识：G1_n
单元名称：启动子及遗传基因转录起始过程
单元事实描述：

(1) 启动子。

定义：综合各种文献中的定义，最后确定。如，聚合酶结合的位点被称为启动子（promoter）。

启动子在遗传基因中的位置。……，如居于上游的一个片段。

启动子的结构。

真核生物启动子。……

细菌启动子，如图 3.10 所示。有一个共有区域，其中心位于转录起始点上游约 10bp 处，长度为 6~7bp，该序列常称为 -10 框(-10box)，还存在另一个短序列，其中心位于转录起始点上游约 35bp 处，被称为 -35 框(-35 box)。……

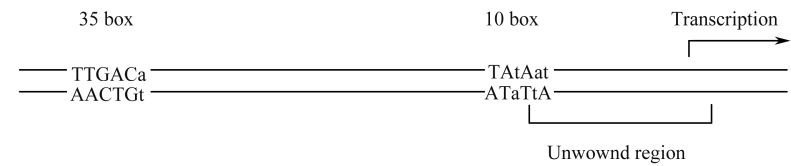


图 3.10 细菌启动子结构

(来源：《分子生物学（原书第五版）》，第 121 页)

核心启动子元件。……，例如，图 3.11 所示 *rrnB* P1 启动子中的核心启动子元件。

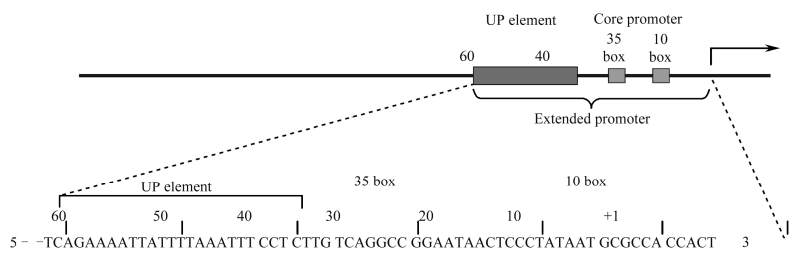


图 3.11 *rrnB* P1 启动子

（来源：《分子生物学（原书第五版）》，第 121 页）

启动子的分子结构。……

启动子的处理功能。

启动子的感知功能。

启动子的连接功能。

与 RNA 聚合酶的连接。

聚合酶 I、II、III，……

沉默子。……

增强子。……

……

（2）转录起始过程。

转录起始过程。如图 3.12 所示。(a) RNA 聚合酶结合在 DNA 上形成闭合启动子复合物；(b) σ 因子促使聚合酶由闭合启动子复合物转变为开放启动子复合物；(c) 聚合酶合成 9~10nt 的初生 RNA 产物，图左侧为一些中断转录物；(d) 聚合酶的启动子清除，进入转录延伸阶段， σ 因子在此时或进入延伸阶段以后被释放。

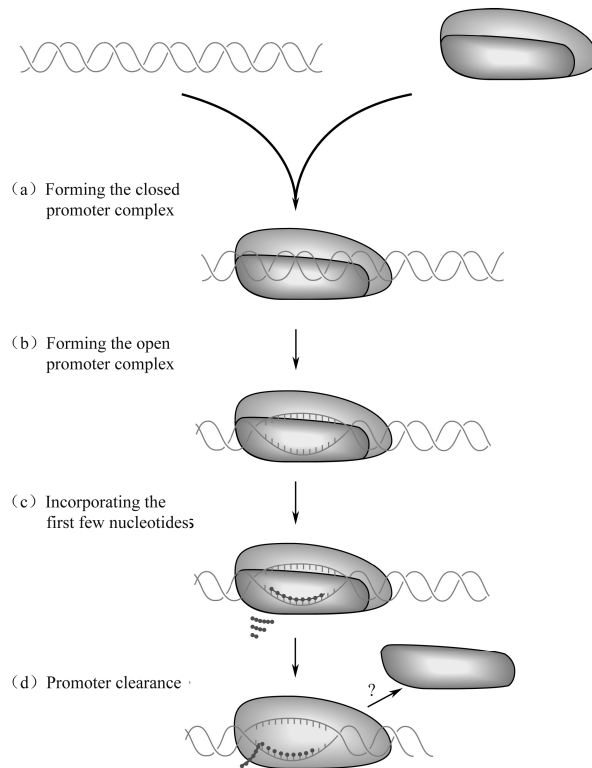


图 3.12 转录起始阶段

(来源:《分子生物学(原书第五版)》,第 123 页)

σ 因子的作用。……

启动子处 DNA 的局部解链。……

处理功能。……

感知功能。……

连接功能。……

这里要指出的是,对于任何活着的生物体,生命过程中的启动子和转录起始过程都是一个完备的单元。假定我们定义一个具体生物体的启动子和转录起始过程的信息单元(也是信息集合)为 $G1_n$,则 $G1_n$ 既可以表示这一信息

单元，也可以表示该生物体对应的自有态信息集合，还可以表示相应自在态信息集合，更可以代表这个客观存在。而记录信息是认知进展的结果，作为自有态的认知信息，既可能是记录信息的映射，也可能是记录信息产生的源泉。对启动子和转录起始过程这个假定的信息单元从自在-自有-记录的分析，证明了信息结构完备的参照系是客观存在本身。

本节没有使用严格的规则 and 标准来构造信息单元和信息结构，这是因为本书不是工程性的著作，去指导如何实现这样的信息结构并构建一个相应的系统。本书是理论性的著作，旨在通过三个例示，说明记录信息的二维静态，跨不同信息形态属性为主和功能为主的客观存在或信息集合的显性信息结构是存在的、可以构建的，并可以不断完善，局部走向完备。

事实上，对上述三个例子，任何具备构建一个特定的信息单元或信息结构的基本知识的人，都可以通过互联网或其他途径收集相关的记录信息，制定相应的规则 and 标准，并按确定的标准和规则不断增加、完善，就可以做出一个达到当今认知水平的信息单元或信息结构。当然，如果需要在信息域共享，则需要按照该信息域共享的规则 and 标准来制定局部的规则 and 标准。进一步，构造或完善一个信息单元或信息结构并不需要只有一个人或一组人完成，而是可以通过互联网，如同维基百科一样由感兴趣、同意按规则参与的互联网用户群体完成。

3.5 信息结构的标识与函数

从 3.3 和 3.4 的讨论，我们看到信息结构由一个个组件、单元构成，动态性强、维度惊人、构件数量巨大。无论是事实、属性的描述还是功能的实现；无论是信息结构完备的过程还是完成任务的过程；无论是载体和外壳的处理还是语义的使用、增长，都需要一个标识体系来实现有效辨识、处理，都需要一个函数体系来应对信息结构形成、使用和发展过程中的复杂变化。本节讨论信息架结构中的标识与函数。

3.5.1 标识的目的和类型

定义 3.23: 信息结构的标识。信息结构的标识是指在信息单元及单元的组件中用以指引内部和外部功能实现的代码或外壳。表 3.1, 表 3.13 及表 3.20 等信息结构形式表述中均有标识列, 标识是信息结构的重要构成部分。

信息结构标识有两个最主要的作用, 一是表示一个信息单元或组件在一个或一组信息结构中的位置, 也可称为地址标识。这类标识有利于信息结构的形式化过程, 更使各类处理更加方便、快捷、有效地进行。这类标识要求唯一、确定、简洁。二是表示一个单元或组件的性质, 指代一个单元或组件的内容, 也可称为语义标识。这类标识使信息单元或信息结构执行内部或外部任务时易以操作。两者比较, 前者更具有形式的性质, 后者更具有语义的性质。

标识使用两类符号, 一类是代码性的, 按照该信息结构确定的规则或标准, 用数字与/或字母表示; 另一类是含义性的, 根据所代表的组件或单元的性质, 包括功能和属性, 通过标识表示出来。一般说来, 用于表示空间和时间的, 确定物理空间位置的适合选用第一类; 用于表示功能或语义关系的适合选用第二类。

遗传信息和认知信息的信息结构同样存在标识。如遗传转录起始的启动子, 既有地址标识, 又有功能标识, 只是通过化学成分及细胞间的顺序来表达; 而视觉系统中视网膜中的视杆细胞、视锥细胞、双极细胞、水平细胞、无长突细胞同样既有地址标识又有功能标识。

上述两类标识符号均可以使用函数, 适用于标识符号串在承担不同的任务时需要改变标识内容的场景。本章 3.4 中大部分标识使用了函数, 这是因为这几个例子都具有过程性特点, 函数对保证结构的稳定性和处理的快捷性具有更好的适应性。

函数主要承担指代的作用, 为信息单元和信息结构的自我成长、表征客体的动态变化、承担任务和场景的不同奠定了标识基础。如地址函数, 可以适应地址各种不同因素的调整; 描述函数, 可以保证结构的稳定, 格式的一致; 功能函数, 可以根据实际任务定义功能。

3.5.2 标识的规范和体系

根据适用的范围，标识的规范有两类，一是关于所有信息结构统一的标识规范，适用于所有信息结构；二是特定信息结构的标识规范，适用于界定范围内的信息结构。

从理论上讲，第一类规范十分重要，有很多理由，如有利于信息的处理和利用，有利于提高信息结构显性的工作效率，等等。其实化学反应不因为化学元素周期表没有研究出来而停止；没有统一的商品代码，国际贸易规模庞大、交易过程顺畅；物质世界的运动和社会活动不因为没有统一的标准而止步。由于适应所有信息结构的规范十分复杂，工作量巨大且困难重重，更由于不同态的信息结构规范实际上没有可能同一，差异大的不同信息域的规范也难以统一，信息结构标识的规范应该是一个从小到大，从具体到抽象的发展模式。其实，标识规范的发展规律是归纳，而不是演绎。实践发展到什么水平，规范跟进到什么阶段是规律的体现。

一个信息结构的标识规范应包括以下几个方面内容，一是每个标识体的结构，二是标识符号的一致性，三是标识函数的体系和结构，四是统一的标识体系。

定义 3.24：一个信息单元或单元中一个组件的所有标识段构成一个标识体。一个标识段是其中标识体中的一个标识，代表一个特定的客体，如地址段、属性描述段、功能描述段等。在一个并存多个信息结构的复杂系统中，一个信息结构也需要一个标识体。一个标识体的结构就是该标识体所有标识段的序列。同一的序列式标识规范的第一步。形如图 3.13：

单元标识	地址标识	属性描述标识段1	……	功能描述标识段1	……	其他描述段标识
------	------	----------	----	----------	----	---------

图 3.13 标识体结构例示

不同的标识符号表述不同的标识含义，同一信息结构中不同体同类标识符号应保持一致。保持一致由相应的规范实现。规范应明确每一个标识段的长度，使用符号的规定，每个标识的内部结构，结构的含义。表 3.12，表 3.13 及类似结构表述都满足这一要求。如果使用函数作为标识符号，则要求满足

对函数表达式的规范。

由于信息单元和信息结构的生长性、动态性，决定了大部分标识段的标识符号需要采用标识函数。规范标识函数一是要规定同类函数表达形式的一致，不同类函数表达形式不能产生矛盾，全部类型的函数表达符号形成一个规范的逻辑体系。

标识体的结构及其规范、标识符号及其一致性规范、标识函数的形式规范和体系合在一起，构成了一个信息结构的标识体系。

需要指出的是，这里讨论的是一个具体的信息结构的标识体及其规范，而不是抽象的一般的信息结构标识体系及其规范。

推论 3.16：一个具体的信息结构，只要已经形成，就能够给出其合理的标识体系和规范标准；只要该信息结构能发展和完备，其标识体系和规范一定能相应发展和完备。这是不证自明的。信息结构能够发展和完备，说明利用该结构的主体具备理解和利用这个结构的能力，而标识与规范是理解和利用的一项基础性能力的体现。

3.5.3 函数的类型和特征

正如前面的分析，函数在信息结构中具有特别重要的作用，是信息结构动态反映其表征的客观实在的变化，信息结构自身发展完善各项功能实现的关键环节，本节对信息结构中的函数作进一步的分析。

信息结构中主要有三类函数，一是地址函数，适应信息结构中地址的动态性特征；二是属性描述函数，适应信息单元中属性的动态性、共享性和单元结构稳定性的需要；三是功能函数，满足功能的变化、共用、实现等需求。

地址函数用来描述间接地址。直接地址是指向一个固定的值，间接地址通过函数指向另一个地址，这个地址中的值是随时间或任务、场景变化的。地址函数通常比较简单，一种是指向一个固定地址，获取存储在其中的当前任务适用的地址；另一种是根据场景或任务规定的方式计算或动态赋值。这些地址的变化规则均来自信息结构对地址确定的定义或算法。

属性描述函数用来处理所描述事实、属性的动态变化、共用，有时为处

理的效率、结构的稳定而通过函数转移描述的物理位置。这三类属性描述函数产生的原因在信息结构中广泛存在，需要用不同的方法处理。对于事实或属性的动态变化，有三个主要原因，一是在信息结构自我完备过程中对描述的修正，二是信息结构表征的客观存在随时间而变化，三是任务或场景的变化，需要动态重组信息单元或单元中的组件，而这种变化的结果是否改变原来的描述体，是不定的。对于前两种情况，可以不用函数，信息结构的描述调整功能可以处理到既定的表述地址中；也可以用函数调用相应处理功能，存放在一个专用的位置中。对于变化的单元或内容比较少，频率比较快的适合用函数方式。对于第三种情况，一般适用函数表述，待任务结束后通过判别类功能函数确定如何调整。

功能单元和功能函数一起完成信息结构的自我完备、应对作为存在依据的外部变化以及完成信息结构或单元承担的任务。在信息结构或信息单元功能实现过程中，信息结构中的功能单元和信息单元中的功能组件承担功能调用和功能实现过程的控制，功能函数则承担功能实现的具体逻辑、算法、模型。

3.3 界定了四类功能：基本功能、逻辑功能、建构功能、任务功能。每一类又可分解为若干个子类，其中的大部分需要通过功能函数来实现被赋予的功能，下面选择几个典型的进行讨论，雷同的就一一展开了。需要再次说明的是，功能函数如同标识体系一样，是针对特定信息结构的，具体的，而不是泛指各种信息结构，抽象的。

感知功能是信息结构自我完备和完成任务的起点，泛在于信息结构及所有信息单元。一个信息结构的感知功能不是其表征的客观存在的物理感知，而是其感知功能的信息实现。因此，感知功能需要对语义、物理、场景三个维度，内部的和外部两个方面的所有相关变化能够感知，但在物理意义上都是符号或外壳的连接，只是需要在恰当的时候、恰当的方式将恰当的对象连接起来。感知功能与连接功能的区分在于感知是获取信息，连接是传输信息。因此，感知功能的函数是在信息单元和信息结构两个层次主动感知与单元或结构相关的变化。实现这个功能，一个信息结构的感知功能函数由两个部分组成：一是信息结构中独立的感知功能单元中确定的感知函数及其算法，二

是各个信息单元中的感知函数及个别特殊算法。一个信息结构主动感知函数的形成一般有五个步骤：一是梳理该信息结构全部信息单元的感知需求，二是对这些需求按算法特点归类，三是确定算法，四是确定各信息单元感知功能部分的描述和标识方式，非通用的算法本身归入该信息单元，五是确定信息结构中的感知功能单元，将信息结构单元间及与信息结构外部的感知功能实现路径及各种通用算法（本信息结构多个信息单元需要使用的）详细地归纳描述道这个或这些单元中。感知算法的特征是触发主动感知的函数的形成，包括阈值和对象，即达到什么标准即触发与什么对象的感知。还要具备与连接功能、处理功能、描述功能、逻辑功能、学习功能、判断功能的连接。

在 3.3 的讨论中已经明确，在一个信息结构中，连接功能是最复杂的功能，一般说来，一个信息结构存在多种连接功能函数和算法，而又主要体现在连接的确定和赋予关系性质。相对于现实世界形形色色的关系，一个具体的信息结构关系是有限的、基本稳定的。一个信息结构连接功能函数的构成、构建的步骤与其他功能函数的相关性与感知函数一样。连接函数的算法特征主要是针对一个个具体关系的确定和本信息结构的特征，一个个具体化。如遗传关系，究竟是时间相继，还是语义相继；邻接关系，究竟是空间位置还是上下文；伴生关系，是物理位置还是语义连带，等等。

就算法特征而言，处理功能函数最复杂。在信息结构中，处理不是计算机信息系统对信息的处理，而是指各种功能通过特定的算法实现。在一个信息结构中只要是函数，就一定有处理的需求，处理功能函数要针对每一个处理需求，算法也要满足所有需求的功能实现。同感知功能一样，许多单元共同需要的处理算法设置专门的算法单元，在相应的单元或调用过程指向需要的算法单元。只有个别或几个单元使用的算法归之于本单元中。尽管任何有一定信息单元数量的信息结构，应该有数量不少的处理功能函数和算法，但复杂程度不高，因为都是已经实现的计算的形式化。

逻辑功能和任务功能，包括归纳、演绎、类比、判断、匹配学习、决策、行为等。根据定义，一个信息结构是一个客观存在的映射，因此，在信息结构中实现上述功能，是该信息结构所表证的客体已经实现的功能，在客体动态发展中，或者在信息结构同步完成客体进行中的任务时，作为信息结构不

应该超越其已知逻辑功能的责任。信息结构不是智能系统，只是为智能系统（生物体或非生物体）实现提供了更加有效的基础。当然，在信息结构的建构和完备过程中，需要运用一些逻辑功能和任务功能，都是可形式化的有限复杂度算法。

在这一系列需要逻辑功能的功能函数中，需要专门关注的是建构中的归类功能，特别是一个记录信息集合建构过程中的通过语义特征函数分解和归类的算法。例如如果我们把 3.4.1 的街景图作为同地点连续 1 个月视频信息集合的信息结构的一个单元，我们如何把每秒 20 帧记，共 51,840,000 个单元中的汽车识别出来，行人识别出来，驾驶员和乘客识别出来，违反交通规则的机动车、非机动车、行人识别出来，树和楼的变化识别出来，成为一个按特定要求构成的时间系列信息结构，一个个单元都由人来识别，工作量太大，现实生活中没有特别的需求，是不可能这样操作的，必须有相应的特征识别函数先行区分，必要时，才通过人机交互，提高可信度。

即使从上述这个简单的例子看，特征函数从整体上看是一个庞大的家族，构成复杂，形式多样，但具体分析，不同的类型均有相当的研究基础，不同的需求实际上只需要少量的特征函数。对于音频信息，语音识别的研究和工程性实现，无论是人的声纹还是不同语言字词的声音特征，都有大量可用特征函数。对于视频信息，20 多年多维数据结构和算法研究、机器视觉研究、模式识别中的有关部分积累了很多成果，而图片图像信息，除了视频信息的研究成果外，还可以利用计算机辅助技术和图像树立技术的成果。对于文本信息，除了人工智能，特别是语义网络和机器翻译的研究成果。传递函数，即所描述的对象中存在物理和语义的传递关系。空间传递函数和事件传递函数是其主要类型。空间传递函数如人、脸、眼睛、眼帘，动物、猫、白猫、猫眼，均是在物理和语义上存在传递关系；时间传递函数如一辆车的运动轨迹，一个摄像头监视区域的时间系列信息，特定声纹存在的延续过程等，同样在物理和语义上存在传递关系。传递函数既需要对其中一个具体对象作为单元构造特征函数，还需要通过传递函数来减少建库和应用的复杂性。对于音视频信息，传递函数的质量，是存储和利用有效性的关键函数。

用于建构的各种特征函数相关的结构和算法，在各类研究和实践中积累

了大量成果。如关于多维点数据的区域树、优先搜索树、四叉树、k-d 树、p-k 树、桶方法等。关于物体图像的非正交块、BSP 树、K-结构、分离链，基于覆盖分裂的覆盖域树、松散四叉树等。关于密度分裂的数字地形模型和数字高程模型、PMR 四叉树等。关于图形边界表示的线四叉树、边四叉树、MX 四叉树和 MX 八叉树、自适应采样距离场、扇区树、锥形树等。关于物体边界表示的条带树、弧线树、BSPR、棱柱树及细化提升和抽取简化等。关于表面边界表示的层级 Delaunay 三角划分、限制四叉树等。关于差别的压缩方法，如形成编码、链码、顶点表示等。关于区间及小矩形表示的线段树、区间树、优先搜索树、MX-CIF 四叉树、平面扫描算法、截窗查询、角缝合法、平衡四维 k-d 树、网格文件锥形区域查找法、LSD 树、二维表示法等。多维数据的结构和算法。如关于最佳优先最近邻查找的相似度搜索、拾取查询、向量量化、快速最近邻分类器、A*算法、轮廓查询、面扫描、有向权重图、增量欧氏限制最近邻法、深度优先 k-最近邻查找、剪枝规则、MAXNEARESTDIST 次序等。多维索引的结构和方法。如 X-树、包围球法、TV 树、os-树，基于距离的索引法，如球划分法、广义超平面划分法、M-树、sa_树、kNN 图、距离矩阵法等。机器视觉研究中相关的数据结构和特征函数相关算法。如视觉图像特征信息提取中，用于图像边缘与平滑的高斯卷积模板，关于边缘检测的 Robeits 边缘算子、Prewitt 边缘算子、拉普拉斯算子、log 算子、Hessian 矩阵法、曲面拟合法，用于角点探测和形状特征分析的 K_R 法、Harris 角点探测、邻点梯度差值法、椭圆拟合法、Hough 变换、广义 Hough 变换等。再如图像融合中的变形轮廓线特征函数相关的算法：基于 Lagrangian 动力学原理的微分耦合动态轮廓线、B 样条形状空间、自适应卡尔曼滤波、离散小波框架算法、图像对比度金字塔等，运动模糊图像标识相关的维诺滤波算法、频域法、钻石搜索法等。在模式识别的研究中同样积累了很多方法可用于构造语义特征函数。化学气味识别、药物分子识别、图像理解、人脸识别、表情识别、手势识别、语音识别、说话人识别、杂草识别等。

其实，在任何一个具体的信息结构中，只需要利用上面的几个算法，甚至有些信息结构的建构功能函数，比上面所述的任何一个算法都要简单，列举这些成果是为了确定一点，任何信息结构，不管其数量多大，逻辑复杂性

多强，总是可以达到建构或完备时智能体的认知水平的。

对于遗传和认知等自有态信息的信息结构，其标识和函数是客观存在的，正是这些标识和函数才使拥有和使用这些客体的主体实现相应的功能。但这是在隐性结构的意义上实现的，实现这些功能不需要显性结构，也就不需要将这些标识和函数可以为第三方理解的形式化，显性结构是信息自我发展和智能进化的需要，将在 3.7 和第 4 章进一步讨论。

3.6 信息结构的一般模型

按照定义，信息结构是一个给定客观存在语义和功能及其内在关系的逻辑体系。而一个具有一定功能的客观存在，通常十分复杂。不论是从多维度的语义关系，还是多场景的应用可能；处理、连接、感知的多中心，还是标识、描述的不定值；不同环境下的多尺度或不定量纲都展示了这一特征。信息结构的一般模型需要对这些属性规范做出回答。上两节分别剖析了信息结构的基础信息单元，介绍了几类典型信息结构的具体形态特征。本节系统描述信息结构的一般模型。

3.6.1 信息结构的构成

满足信息结构对应的客观世界的复杂要求，又使结构尽可能简洁，信息结构的构成有两大部分，即信息单元和单元间的连接。

如前所述，信息单元分两类，一类是属性单元，一类是功能单元。

属性单元分不同层次，最高层次是对整个信息结构的整体属性描述，最低层次是最小信息单元，中间层次根据信息结构的属性和任务复杂性来确定。

功能单元一般为两个层次，基础层在信息单元中，每个信息单元拥有该单元存在和发展的必要功能，描述、感知、连接、处理。信息结构中独立的功能单元是实现功能的主体。3.5 中确定的四类功能的每一种具体功能均有一个或一组功能单元满足功能的实现。一个具体的功能至少由三个部分构成，

一是分解的可执行的子功能，每个子功能有一个信息单元具体描述；二是功能实现的过程，从触发到各个子功能的逻辑顺序；三是与各个信息单元的相应功能的连接与协同。一个信息结构如果特别复杂，在不同的子集中存在特殊的功能，功能单元也可以分层。

每个信息单元，不管是属性单元还是功能单元，其结构均包括属性、事实的描述和功能的描述，在整个信息结构中，通过标识体系将单元和功能在语义和执行两个方向形成整体，为辨识和执行提供接口。

连接是信息结构成为整体和执行功能的基础条件。观察遗传基因及认知神经系统的组织结构、细胞间的关系，连接在其中占有特别重要的位置。在过去的实践中，记录态信息除少数系统能够外，基本上没有构建实现信息集合功能的全部连接。

连接是信息结构功能的组成部分但在各项功能中具统领的作用。一个形式化描述体系，之所以能称之为信息结构，是通过连接来体现的。保证整个信息结构个个组成部分充分、有效连接是信息结构的根本任务之一，3.6.2 的模型中体现了这个原则。

3.6.2 信息结构的一般模型

利用 3.6.1 的分析结构，可以勾画出信息结构的一般模型。图 3.14 是信息结构一般模型的简要图示。

图 3.14 中，最中间的正方体表示该信息结构全部共同功能单元，通过全联通的环形通道与所有的信息单元连接，并通过每个信息单元连接功能实现共同功能与信息单元中的具体功能协同。最外边的环形是该信息结构与外部感知、连接的通道，功能单元通过内环形通道实现与外部的交互，各个信息单元通过自身的连接和感知功能实现与外部的交互。各功能单元间的交互也通过内环形通道实现。

图 3.14 没有表述信息单元间，特别是属性单元间的层次关系，而在信息结构中层次关系对描述和功能实现都有重要意义。图 3.15 是属性单元层次关系的图示。

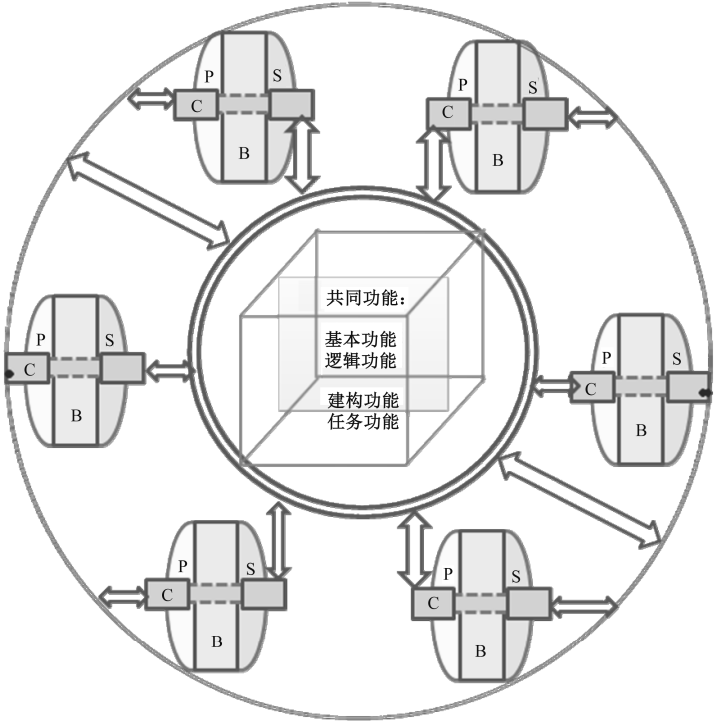


图 3.14 信息结构一般模型示意图

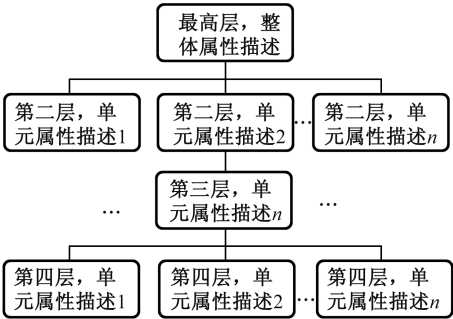


图 3.15 属性单元层次关系示意图

一个信息结构生成什么样的层次结构,基于内在的语义关系和任务特征。如图 3.4 中关于视网膜的例子中,若设视杆细胞为最小信息单元,则应该是第四层,相应第三层是外核层。第二层是视网膜,第一层是视觉系统这一信息结构的整体描述。

一个信息结构中,每个功能应该具有独立的可执行能力,图 3.10 只是逻辑上可推论具有这一功能,图 3.16 以表 3.11 中关于视杆细胞感光功能来说明如何基于图 3.14 实现具体的功能。

如图 3.16 所示,光线经由单元的外部感知圈触发感光功能单元,感光功能单元通过内连接启动视紫红质单元的感知、连接和处理功能组件,再调动视黄醒单元的感知、连接、处理功能组件,再相继继续调动视蛋白和感受器单元的相应功能组件,诱发视杆细胞产生感受器电位,引起其下游的双极细胞产生电位改变,进一步引起神经节细胞放电频率发生变化,逐级传递至视皮层产生视觉。这里双极细胞单元则是与视杆细胞同一层次的另一个信息单元了。

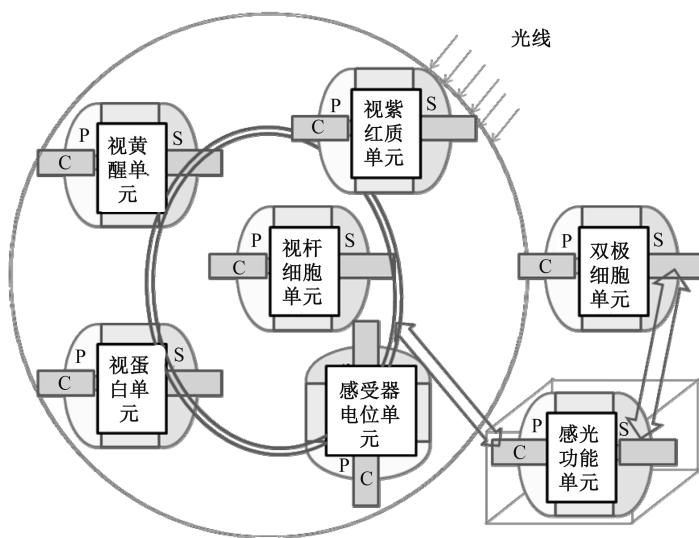


图 3.16 视杆细胞感光功能的实现路径示意图

3.6.3 最简信息结构与冗余

在一个信息域的信息结构体系中，必然存在最简信息结构和具有不同冗余度的信息结构，但产生这两种类型的信息结构的原因并不相同。冗余的产生主要由于客观存在的多维度特征与相当部分信息结构显性单维度目的的根本性矛盾，最简信息结构的产生源自信息结构体系发展的需要，需要一个最完备又在语义上不重复的最简信息结构来作为其他信息结构形成和发展的基准。

定义 3.25：最简显性信息结构和最简完备显性信息结构。最简显性信息结构是指在一个给定的范畴内，一个信息结构中所有构成部分都不存在冗余。最简完备显性信息结构是指给定范畴中，有一个信息结构既是最简的又是完备的。这一定义有三层最主要的含义，首先，最简结构是一个范畴的概念，在一个给定的范畴中才能比较，才能为同一范畴中其他信息结构显性提供基准或参考；相对于信息结构，这个范畴可以是一个主题域（或任务域），信息域甚至是全域；相对于信息单元，主要界定在一个信息结构中。其次，是最简显性或完备信息结构一般不是自发在任务驱动的信息结构显性过程中形成，而是为了形成完备结构的目的而专门进行的。最后，必须是完备的或最简的，才能达到成为其他信息结构显性的基准或参考的目的。诸如上下层之间的重叠部分、多场景中的重复部分、功能实现过程中的交叉部分，等等都用指代的方式，通过地址或函数实现调用。

推论 3.17：在一个信息结构或信息单元中存在冗余构件的就是冗余信息结构或冗余信息单元。冗余信息单元是指信息单元中存在冗余的描述或功能。冗余信息结构则既包括存在冗余信息单元，也包括信息单元间的冗余。

冗余的产生是必然的，因为有三个不可避免的因素决定了冗余的产生：一是本节一开始指出的，信息结构对应的客观存在的多维特征与一些信息结构显性的目的只要求一维或少数几个维度，多个相同维度的任务必然产生结构显性的冗余；二是任务的需要，不少因任务而产生的信息结构，冗余易于理解、易于处理；三是简化会增加经费和人力的投入，有时也需要更长的时间。

冗余产生在信息结构的各个层面。一个信息单元内不同的点可以存在重复，一个信息结构内的信息单元可以存在重复，同一信息结构不同信息单元的类似的点可以存在重复，不同信息结构间的信息单元和信息单元中的点可以存在重复，同一主题域不同信息结构及不同信息结构中的构成部分可以存在重复，同一信息域的不同主题域间的信息结构可以存在重复。

冗余不仅可以产生在信息结构的不同层面，冗余的程度也存在很大差别。衡量冗余程度的指标成为冗余度。

定义 3.26: 信息结构冗余度。信息结构冗余度是指一个显性信息结构或结构体系中，相同信息点的重复程度，是一个整数值。不管产生在什么层级，冗余度同一以最基本的构成信息单元的信息点为计量单位，一个信息点的重复的冗余度为 1，以此类推。

定义 3.27: 是绝对冗余度，但不能衡量一个信息结构中冗余的比例有多大。

推论 3.18: 信息结构相对冗余度。设一个信息单元或一个信息结构共有 N 个信息点，其中 M 个是重复的，则该信息单元或信息结构的相对冗余度是 N/M 。因此，相对冗余度是衡量一个信息单元或一个信息结构中冗余的比例。

信息结构语义多维度、任务多场景、处理多中心等特征，必然存在是否保留冗余，如何保留等需要回答的问题。

实际建构过程中，通常是保留适当的冗余，而不去追求最简结构，根据实际的需求和可能走显示的路径。最简结构增加了算法的复杂性或功能实现过程的复杂性，增加了整个结构中核心单元的工作量，增加了计算复杂性。但过多保留冗余，导致数据量过大，也会增加了存储和处理成本和复杂性。特别是对于特定应用，只对可能使用的部分作为信息结构的对象，对使用中发挥作用很小的部分忽略或不进行动态结构化的方式处理。为特定目的服务的视频，大都具有这个特征，如记录机动车是否闯红灯的视频，需要把过红灯时违规的机动车筛选出来，就要把机动车过路口停止线的几秒钟与灯光匹配结构化，其他时间的视频就可以省略了，而且对这几秒钟的视频也侧重可判断是否闯红灯的部分结构化，其他部分可以简化或省略。

有的冗余与当前的应用没有关系，甚至在信息集合存续期间都可能不会利用，但还是保留了冗余，例如，遗传信息中的暗基因、非编码基因，以及

认知信息中记下来之后从没有再激活过的信息，虽偶有激活，但长期不用，最终湮没的信息。这种冗余存在的原因主要是两点，一是存在过，二是可能在一些非正常的过程中会使用。记录信息中也大量存在这样的信息，这种冗余是信息结构为增长和调整留下的余地。

对于最简和冗余、冗余幅度之间的取舍，有些是比较容易决定的，如前两类，结构或数量带来的计算复杂性、存储成本等都是可以定量评价的，对于第三类冗余，只有通过长期总结发展实践才可以归纳出可供判断决策的准则。

3.7 信息结构的可用性

信息运动的历程是从自在到自有记录，从隐性结构到显性结构，从信息结构的不完备到完备。这个过程的载体或核心概念是信息结构，动力则是信息的可用性。本节讨论信息的可用性和信息结构的可用性。

3.7.1 定义

定义 3.28：信息的可用性。信息的可用性是指主体对拥有或得到的信息能否使用及使用的程度。定义中的主体是指具有感知、操纵、获取信息功能的生物体或非生物体。定义中的信息包括全部三种存在形态的信息。定义中的能否使用是指能否为主体的行为或目的产生影响，可用是个目的性概念。

定义 3.29：信息结构的可用性。信息结构的可用性是指主体对拥有或得到的信息结构能否使用及使用的程度。定义中的主体及能否利用的解释同定义中的信息结构包括隐性结构和显性结构。

按上述两个定义，信息的可用性和信息结构的可用性存在大量的交叉，一般情况下，主体使用的是结构化的信息，区别在于隐性的还是显性的，所以下面只讨论信息结构的可用性。

3.7.2 信息结构可用性的一般过程和构成要素

信息结构利用的一般过程是：相应主体根据当前需求，利用拥有的或可获得的信息结构，达到当前需求目的的过程。这一过程中，主体和信息结构间的关系如图 3.17 所示。

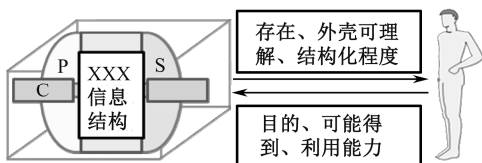


图 3.17 主体和信息结构的关系图示

如图 3.17 所示，信息结构侧可用性构成要素是：一是存在与主体当前需求一致的信息结构，二是信息结构使用的外壳与相应主体的一致，三是信息结构与相应利用过程存在可利用的通道。而利用主体侧可用性的构成要素是：一是需求，即存在利用的动力，二是具备利用的能力，如能得到、能理解、能处理等，三是具备相应的环境，如过程的可控性、时间和空间等因素。构成要素也是必要条件，如果这些条件缺失，信息结构就失去了可用性。

3.7.3 隐性结构的可用性和局限性

任何信息集合都是客观存在的一种表征，客观存在包含的含义及其相互关系就是该信息集合的内在结构。这样的结构有的是隐性、有的是显性，隐性信息结构是主体。一般说来，隐性信息结构是客观存在的映射，客观存在决定了隐性结构，但也有例外。如人在进行艺术性创作，不管是美术、文学、音乐，作品这个客体是创作者大脑中的存在转换到这个作品中，但作品的客体完成后，创作者大脑中可能不保存或不能记忆全部细节，作品这个第二客观存在反而是最完整地保留了全部内涵，但这个内涵的解释依赖第一客观存在。这种依赖第一客观存在解释，或者说这样的信息集合的隐性信息结构到

显性信息结构的转换依赖第一客观存在的现象不仅在艺术领域，在其他学术领域也存在。

隐性信息结构是可用的，下面分别对自在态、自有态的遗传和认知、记录态信息隐性结构的可用性及局限性进行讨论。

3.7.3.1 自在态信息隐性结构的有用性

自在态信息都是以隐性结构存在，存在是被动的，与目的无关，但因存在而有用。自在态信息结构为自有态、记录态信息结构所用，利用基于可感知、可记录。

自在态信息隐形结构是自有态、记录态信息发生发展的基础，是显性结构完备度的参照系（见本书前面相关章节）。

自在态信息隐型结构是感知之源。生物体感知外界环境变化，并采取相应的调适，非生物信息系统感知控制对象的状态和变化，都以感知对象的特定时空状态为基础。对于感知主体，交换的是能量，获得的是信息。从物理世界看，感知是物质运动，从信息世界看，感知是信息运动，因为不管是何种感知，需要得到的是信息。

生物体对外部环境的感知，是自在态信息转换为自有态信息，进而使生物对外部环境的调适做出反应。这类作用的产生是有条件的。生物体必须拥有对外部环境变化相应的感知细胞、能量转换为化学或电信号的功能、信号逐级传输到可以采取调整的部位，并存在这样的调整功能。例如，积温是植物发芽、生长、开花、结果、落叶的主要参数，植物在生长过程中，能感知温度变化，具有传导、累积功能，并有相应的基因组根据信号启动或关闭。而动物由于有神经系统而感知功能更加成熟。例如，在人类的皮肤上有专门的细胞感知冷热，手的表面冷点有 1~5 个/平方厘米，热点有 0.4 个/平方厘米，这些冷热点分布于皮肤表面 0.15~0.17 毫米处，并逐级连接到中枢神经，形成温度的感觉。

生物体的感知是兼具生命、物理、信息三种形态的运动，应该先从三个维度独立解释，然后在综合。从生命运动的角度看是生物生存和发展的一种手段，是生命过程的一个环节，是生命体与外界联系的起点；从物理运动的

角度看,是一种能量转换过程;而从信息运动的角度看,则是自在态信息向自有态信息的转换过程,是自有态隐性信息结构动态变化或完善的过程。从三个不同维度看同一个过程是全面认识这个过程的起点。没有外部物理运动的客观存在,即温度、光照、压力等变化的物理环境,生命体即使有感光、感温、感压力的细胞,感知过程没有起点;物理环境存在,没有生命体将感受到的物理状态转换为电信号或化学信号,并传导到中枢神经系统(高等动物)或相应的调控功能区(没有中枢神经系统的其他生物),自有态隐性信息结构的产生变化,促使生物体对环境变化做出响应,这个过程没有意义;仅有自有态隐性信息结构即时变化,没有对遗传基因突变选择时发挥作用,就不可能实现生物进化的正向选择。因此,从生命的生存和发展看,这个过程发生在生命体和物理存在之间,而发挥作用的是信息的运动。这一信息运动过程的核心是自有态信息隐性结构的因变和积累,前提是生物体感知、自有态信息结构的存在及其与生物体其他器官的协同能力。

自在态隐性信息结构有用性基于能否对自有态和记录态信息结构提供增量。增量有即时增量、周期增量和永久增量。即时增量是指感知信息仅用于主体当前事项的处置。就生物体而言是指即时应对外部环境的变化,如葵花绕着阳光转,而没有留存于自身的隐性信息结构中;就非生物体信息系统而言,是指即时一次性使用的感知信息,如炼铁高炉传感器获取的信息用于实时控制,而不保留记录。周期增量是指保留一段时间后不再保留的感知信息,失去后不复存在。自有态信息是指有效的记忆时间,不同物种,不同刺激强度,保留的时间或周期不同。记录态信息是指保留一个时间段,过了规定的时间即不再保留,如互联网即时通信的信息,路况监控信息等。永久增量是指长期保留于自有态或记录态的信息中。自有态信息是指感知转换为遗传基因,包括基因触发开关的改变,记忆成为与生物体生命同存的长期记忆。记录态信息是指经过或不过转换,感知信息存在于可以长久保存的记录信息载体中。

自在态信息隐性结构对无生命物质是否存在着影响及如何影响在本书的附录讨论。

3.7.3.2 自有态信息隐性结构的可用性

自有态信息结构有三个大类，一是生物体遗传信息，二是生物体认知信息，三是非生物体信息系统中的信息。

生物体遗传信息隐性结构利用是一对一的，基本不变的。即拥有该遗传信息的人可用其隐性信息结构完成生长发育过程，而且，即使在遗传过程中发生突变，生物体拥有的是变化后的新构成。因此，可用的前提就是生存环境的存在，就是生命的存在及遗传信息发挥作用的生命机理存在，即信息传递的通道和控制机制的完整。

生物体认知信息隐性结构的可用性比遗传信息要复杂。它由两个部分构成，一部分是具有信息传递、存储、处理功能的中枢神经系统，即一般意义上的认知信息；另一部分是具有信息感知、传递、反应能力，但基本不具备存储及存储后的处理能力，或不需要中枢神经系统集中处理感知信息然后做出反应，如植物神经系统，也称自主神经系统，处理生理性认知信息，感知内脏，并根据感知的信息做出反应，但不需要大脑的信息处理、决策功能。与遗传信息结构的不变相比，认知信息结构的特点就是时刻在变，生物体不断感知或获取新的信息，信息结构依据新得到的信息不断变化；生物体在不断进行学习，情绪和需要处理的事务也在变，大脑中的信息结构也因此而变。如果说遗传信息决定生物体的生长发育，则认知信息则决定生物体的行为（认知行为和生理行为），这也是两种自有态信息的根本不同。

植物神经信息结构的可用性原理与遗传信息结构相似，前提是相应神经系统从感知、传递到反应的功能完整，生物体生命的存在，但其信息和控制的复杂性比遗传基因信息结构简单。

除植物神经隐性信息结构外，决定其认知信息结构可用性的因素更多。从感知细胞开始到中枢神经及大脑的信息传递和转换通道必须通畅、过程无误，信息合成、存储、处理、反应之间的模式、物理、化学过程与信息过程的高度一致，细胞和神经系统在传输、存储、处理过程中的信息外壳在转换中保持同一性，这是一般生物体的要素和结构。对于已经存在于认知信息，特别是大脑中已存信息的语义外壳一致的记录信息，要在处理和存储过程中，实现不同符号体系的归纳，在大脑中实现不同外壳的融合，并根据记录信息

符号和概念的体系的更新而更新。

可感知、可理解、可传递、可存储、可处理，并在这个过程中保证物理、生命过程与信息过程的一致，不同外壳合成的一致，存储能力、处理能力、不同外壳体系之间的转换能力与生物体行为对信息处理要求一致，这些都是决定认知信息结构可用性的要素。

第三类自有态信息，即非生物体信息系统中感知、获取、存储和处理的信息，且不属于共有信息，只有该系统可以辨识，只供该系统使用的，而不同于其他记录态信息的符号和概念体系的，基本上还没有产生。其最关键的不同是否存在独一无二、非多主体共有、超越现有的符号和概念体系的外壳。目前还没有存在具备这一特征的非生物体信息系统。在以后的发展路径中，独立外壳非共有将是一些系统发展的中间环节，经过一个阶段发展后必然走向共有，其规律就与共有显性结构相同，因此，不在这里专门讨论。

3.7.3.3 记录态信息隐性结构的可用性

记录态信息的隐性结构是指一类已经记录下来的信息集合，它内在地表征了所记录的客观存在的含义及关系，但含义和关系没有利用该信息的主体（人或计算机信息系统）可辨识的外壳表示。这类信息的数量很多，如遥感信息、地震波勘探信息，各类没经过处理的感知信息大体上属于此。

除开少数特例，这类信息的隐性结构，只有通过结构显性才能用。在下一节讨论隐性结构到显性结构的转换并利用问题。

3.7.4 显性结构及其可用性

在信息结构中，隐性结构是主要组成部分，显性结构基于隐性结构，相应的隐形结构不存在，显性结构没有存在的基础。结构显性是信息增长中的一个核心环节，没有结构显性，信息增长过程必然中断，无法实现智能和文明的进化。本节从四个方面讨论显性结构及其可用性。

3.7.4.1 隐性结构与显性结构的关系

所有的客观存在都有对应的信息集合，所有的信息集合都有对应的隐性

信息结构，因此本书有的时候就忽略了信息集合这个中间环节，客观存在直接对应隐性信息结构。因此没有隐性结构就没有显性结构。

不仅显性结构基于隐性结构，而且显性结构所揭示的语义和关系，也需要经由隐性结构的验证，只有隐性结构存在和使用的含义、功能及关系才是有效的。基因组工程是将基因的隐性信息结构显性化。每一个基因的编码及功能实现的描述，都以隐性结构、也就是客观存在为依据，离开了这个依据，就是错误的。

显性结构是隐性结构用多个、有时候是多类主体可理解、可利用的外壳将其中的部分或全部语义及其相互关系表示出来。这个表示不同于人工智能的知识表示。人工智能的知识表示针对一个特定逻辑过程的知识需要，形式上要适合该系统逻辑处理架构及模式，内容上一般不包括功能性的知识知识表示。结构显性针对一个客观存在包含的语义及其关系，形式上在遵循一般表示规则要求的前提下，要适合所表示的客观实在的特征。相对于客观存在，知识表示是信息结构显性的一个子集，有时候是一个很小的子集。

从可用性看，隐性结构有两个基本特征，一是被特定主体所用，二是若要是为其他主体使用，要进行转换或将其隐性结构显性化。而显性结构不同，信息结构从隐性到显性，目的就是使信息变成共有信息，变成多个、甚至多类主体可理解、可用的信息，因此显性结构的可用性是既是单个利用主体的函数，也是多个（类）主体的函数。

3.7.4.2 显性信息结构可用性

根据上一节的分析，显性结构是否可用、可用程度主要体现在与利用主体的匹配性，而且也仅限于记录态信息。

首先是显性结构使用的外壳，是否与利用主体一致，一致性如何。如果不一致，显然不可利用。实际上，任何记录信息结构的显性，都是以一类智能主体适用的符号及概念体系为基础，必然能为与此一致的主体匹配。而不能理解这类符号和概念体系的主体就不能利用了。还有更深层次的问题，即与显性结构使用同一符号和概念体系的主体是不是都能利用，答案是不一定。假设一段关于一个细胞感知的信息，并传递到神经细胞的突触的显性结构，

用细胞壁钠、钾、氯离子通道开关的原理解释，对很多非专业人员可能一头雾水，而用吸管或其他形象比喻可能会恍然大悟。所以仅仅符号和概念体系的一致，并不表示一定存在高的匹配度，高的匹配度依赖于显性结构与利用主体的对相应事务理解的一致性。这也直接引出另外一个问题，即同一客观存在及其信息集合的结构显性并非是唯一的。一是由于需要用不同的外壳表示，二是需要针对不同的使用对象采用不同的描述语言，三是针对不同的使用场景需要采用不同的结构模式。

其次是信息结构表述的客观存在的颗粒度。相对于具体的利用场景，对颗粒度的要求存在差异。如对一段公路交通监控录像中的一辆轿车的信息结构，如果是要确定的违反交通规则的车，就要能分辨车型、车号、车主要外表特征；如果是要确定是否为一个案件的逃犯人，就要能分辨司机和乘车人，能分辨司机和乘车人的脸部特征、服装，分析脸部特征即要与已经掌握的人脸特征识别函数一致。颗粒度不仅与利用时的场景相关，也与利用主体自有认知信息结构的颗粒度模式相关。与利用主体相关事件的自有信息结构的匹配度决定着利用的速度和质量。

最后是信息结构表述与对应客观存在的符合度。符合度反映在两个维度，一是正确性，二是系统性。正确性是指表述的结构，不管是描述还是功能，连接还是计算，是否与其指代的客观存在一致。系统性是指该信息结构与任务相关的信息集合，即客观存在的集合是否一致，至少不小于任务相关的信息集合。正确性是从语义逻辑的维度规范了可用性，系统性则从任务场景的维度规范了可用性。

3.7.5 信息结构的显性

显性的信息结构是连接自在态、自有态、记录态信息的中介，是研究信息增长和演进的基础。信息结构显性的动力来自智能体生存的压力和理解未知的渴望。人类产生以来，始终努力将隐性的信息结构显性化。但是否所有的信息集合，或客观存在的隐性信息结构都可以显性化，约束条件是什么，这是本节讨论的主要内容。

动物的中枢神经系统，特别是大脑产生以来，生存的压力和对未知的恐惧，尽力去理解围绕自身的世界，自然界和生命界，并将这种理解，以及利用已经取得的外部世界的认知有利于生存的经验告知伙伴。蜜蜂用肢体语言将蜜源告诉蜂群中的其他成员，站岗放哨的候鸟用鸣叫告知遇到的危险，诸如此类的行为实际上经历了两次信息转换过程，实现了最基本的隐性信息结构显性化。第一次是将自在态的信息转换为自有态信息，即蜜蜂或候鸟感知到了蜜源或入侵风险；第二次是将这一自有态信息用本生物群体共有外壳，肢体语言或声音表述出来，即从自有态信息转换为记录态信息。这个过程也是直接将自有隐性信息结构转换为显性结构的过程，间接将自在态隐性信息结构转换为显性结构。

信息结构显性化过程中最为关键的是人类的语言和文字的产生。语言的产生是人类进化中里程碑式的重大进步，但显然是人类信息能力提升的一个跃迁，是人类的自有态信息向记录态信息迈进的前提，是比鸟叫更加深入、系统的实现自有态隐性结构向显性结构过渡的跃迁。

晚期智人在公元前 10 万年时出现在南非，用了 7 万年时间，在世界范围内取代了所有其他原始人类。实现这一变革的是他们掌握了语言。有了语言，他们可以比其他生物，也包括其他原始人类更加丰富的信息表达方式，而这种表达方式，不仅比其他人类在信息的交流、复杂行动的协同上占据了明显的优势，而且在有了语言这个信息的载体之后，更使知识的积累有了更好的表述方式，能逐步发展出表达抽象思想的概念。正因为这个原因，经过 7 万年的演进之后，到公元前 3 万年这种抽象的思想和概念能以绘画、线描、雕刻的形式展示出来¹。

文字的产生同样经历了漫长的演进过程。语言因何产生？没有科学的证据，但模仿大自然和其他生物的声音，从而改变了声带和喉部的结构，形成更加复杂的发音系统，经历数万年的进化，从简单有含义的发音到形成复杂的语言系统，可能是一种途径。文字产生于语言之后，人类生存环境有所改善，能够有稳定的居住和一定的余暇的时候，把看到的物体和自己的行为记

1 参见《全球通史》第一卷，人类的黎明，美国时代生活编辑部著，赵沛林译，吉林文史出版社，2010 年 1 月，第一章。

录下来的欲望，记录过程中不断企图把认知过程中积累的抽象概念用图形表述的动力，推动着文字的产生。考古研究发掘出来的图像和古文字及相关研究证实了这个过程。从信息结构的角度看，这个过程更加清晰地展示了人类把自在态隐性结构转换并改变自有态信息结构，最终把自有态信息结构转换成记录信息和显性信息结构。这个过程不仅实现了自在态信息到认知信息的转换，还实现了遗传信息的改变。晚期智人从公元前 10 万年以发音基因的初始突变为基础，经过 7 万年的进化，成功地进化到现代人体型²。

语言在自有态认知信息向记录态信息转换跨出了第一步，但在录音技术发明之前，只具有即时共用的特征，还不能历时保存。表达认知信息含义的画及象形文字的产生，使人类第一次有了将认知信息转换成历时性记录态信息，也有了最早的记录态信息显性结构。图 3.18 几张图说明了这种早期的信息形态转换和结构显性。

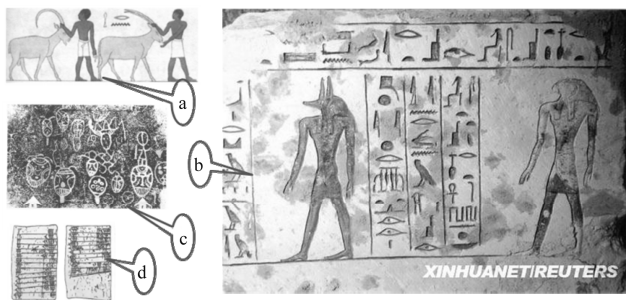


图 3.18 文字从象形到符号的进化

(a 是以画显意，b 是画与象形符号合作表达复杂概念，c 是象形符号的表述，d 是表述抽象概念。来源见参考文献³)

2 同上。

3 图 3.18a 转引自高嵩 高原著，《岩画中的文字和文字中的历史》，宁夏人民出版社，2007 年，第 209 页。b 转引自天山网“埃及发现三千年前贵妇墓”，网址：<http://www.tianshanet.com>。c 同 a，第 97 页。d 引自 James Gleick, *The Information, a History, a Theory, a Flood*, Fourth Estate, London, 2011, p.44。

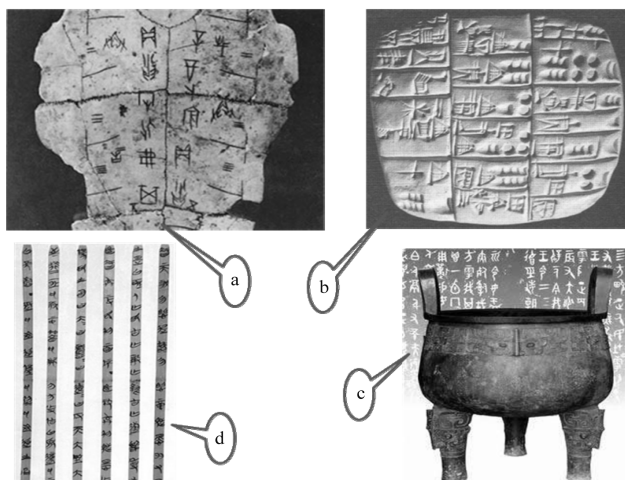


图 3.19 文字及载体的进一步演进

(a 以龟甲为载体, b 以泥板为载体, c 以青铜器为载体, d 以竹板为载体。来源见参考文献4)

如果图 3.13 说明了文字从象形到表意的形声, 从形象表达到符号表达的渐变过程, 那么在公元前 1500 年左右出现在地中海东岸的字母表达, 则是一个革命性的突变。闪族人发明的 22 个没有意义的字母, 每个字母是最小的声音单元, 每个组合 (组合可以是一个字母, 也可以是多个字母) 构成的词表达确定的含义, 这一模式显然有利于学习、有利于表示复杂含义。尽管它的产生比象形文字晚了数千年, 但其易学、灵活、有效的特点, 在很短的时间内走向世界, 成为大多数国家和地区人类表达认知世界的符号体系。而几乎是同时, 数字符号在两河流域初步形成系统, 图 3.12 的 d 就是数字计算的泥板⁵。

从象形到形声、字母, 以及数字体系的产生, 不仅是表述方式和符号及

4 图 3.13a 转引自中国民乐网-华人音乐网-民族乐器网。b 引自 James Gleick, *The Information, a History, a Theory, a Flood, Fourth Estate, London, 2011*, p.42。c 和 d 同 a。

5 参见 James Gleick, *The Information, a History, a Theory, a Flood, Fourth Estate, London, 2011*, chp.2。

概念体系的进步，也为信息隐性结构到显性的转换提供了新的载体，形声和字母构成的文字体系，成了人类将认知信息转换为记录信息的有效工具。

任何一个记录信息都在一定程度上是显性结构，记录过程的逻辑就是这一信息集合显性结构的框架。图 3.12 和图 3.13 中的任何一个记录，都有着明确的逻辑架构，而后代对这些文字的破译或解释，都是一个个图像或象形文字的含义和其中的逻辑结构结合起来进行的。

当一个含有文字的甲骨或泥板放在面前，从信息结构的视图看，刻写这一记录的人是理解这一记录的所有显性或隐性的结构，但这一结构今人不理解了。直接的是所用文字的含义不理解，间接的是文字之间的关系如何区分、记录的空间布局有何含义也不能理解。这个不理解是因为形成记录的人和解释记录的人处于不同的信息域，也就是信息含义的表述使用不同的外壳。但是，我们应该认定，在同一个信息域，至少有一类人是可以理解并解释这一记录下来的信息的，也就是说，这个记录态的信息结构是可以显性化的。

语言和文字的产生推动人类社会加速向前发展，记录的信息就越来越多，语言文字的稳定性也越来越强，上述的不同信息域逐步减少，即使是使用不同的象形文字和字母文字形成的不同信息域，由于处于相同时代，因交流的需要，出现了日益增加的理解两个或两个以上语言的人，使得这种基于外壳的解释或结构显性成为可能。

记录态信息的增多，出现了将许多文献收集在一起的信息集合，集合中的数量达到一定程度，需要信息集合整体的部分结构显性。当西汉汉成帝集天下之书达一万三千余卷时，如不将各卷之间的结构显示出来，利用则十分不便，于是命刘向将书分类、编号，做出内容提要，形成了以书为单位的简单的显性结构。于是人们可以按图索骥，不必在茫茫书海中逐一搜寻。随着文献量的增加，出现了按篇名、主题甚至一些经典著作的字词作索引的方式，显性结构中的标识单元由一本、一卷细化到篇或字词，但单元间含义的内在关系显示仍十分薄弱。

教育的发展、科技文化的重视、记录信息的技术突破，工业革命以来记录态信息快速增长，互联网、传感器进一步加速了这个进程，大数据这个概念的反映了记录态信息高速增长的事实。记录态信息的增长，冲击着信息组

织管理方式，推动了记录信息组织管理技术的发展。首先是数据库技术，以一套严格、有效的规则管理和处理数据库里的信息，其核心就是信息结构的显性。对数据库管理的信息单元进行明确的物理和语义的定义，将这些单元之间的关系在语法（形式和连接方式）、语义（信息单元的定义和相互间关系）、语用（用户模式）三个方面显性地表示了出来。一个或多个数据库的数据模型的显性结构表示的关系也只是这些数据全部隐含关系中一个很小的子集。随着机构或系统信息的增长，这些数据库管理和处理的仅是该机构或该系统拥有信息的一个子集，且对音视频、文档的管理和处理需要更有效的工具，于是以 Hadoop 为代表的所谓大数据管理工具开始出现。随着信息和应用的发展，新的工具将会逐渐适应这一变化。数据库、大数据管理工具及今后产生的新的管理工具，是推动记录态信息结构显性的技术力量。

组织机构，特别是企业，面向应用的信息系统越建越多，系统间的信息如何实现有效分享，如何使这些信息更充分发挥作用，拥有大量信息的机构和企业开始构建覆盖全部信息的统一的信息模型，记录信息结构的显性表示进入了一个新维度。一个组织机构或企业中的全局信息模型以该机构中生成和使用的记录态信息为对象，以使用的最小单位作为表达的基本单元，本机构可预见应用对这些单元相互关系为基础的信息结构显性表达。

以一定的应用目标，组织信息，并将其在应用目的水平上深度结构化，这是人工智能领域知识表示的成果。专家系统的知识库及其表示就是这类显性结构的典型例子：将某一专门领域（通常是十分专门的领域）的相关知识尽可能系统完整地收集并表示出来，不仅包括已经文献化的知识，还包括领域专家的经验 and 诀窍。知识表示的多年努力，演化出了一系列对信息结构显性有重要参考价值的工具和方法，如语义关联和语义网络、脚本、框架、代理等。与目录、索引、信息数据或模型相比，人工智能在给定领域知识表示所揭示的信息结构是最充分的。

但是机构对其所拥有信息的结构显性受工具和应用视野的约束，人工智能的知识表示受领域范围和相应人工智能系统使用的推理工具的约束。在一般的人工智能系统，知识表示从属于逻辑推理过程。在逻辑推理为主还是拥有的知识及其表示为主，在人工智能领域经历了长期的争论，在一些知识（信

息)密集的人工智能系统中,特别是在深度学习和深度分析系统中,将更充分收集信息,并将许多基于推理功能通过对信息单元颗粒度的细化和关系的全面展开来替代。2011年2月14~16日在美国“Jeopardy”电视娱乐节目中战胜两个该节目胜率最高者的“沃森(Watson)”,对节目可能涉及的广泛领域的知识以及在提问时充满机智、诙谐、陷阱的语言环境下,如何实现在极短的时间内从语音识别到事实搜索、知识匹配、组织回答语句,核心是拥有了海量的信息,沃森存储了大量图书、新闻和电影剧本资料、辞海、文选和《世界图书百科全书》(World Book Encyclopedia)等2亿页,数百万份资料,并对这些信息进行了与任务一致的、充分的结构化处理。

沃森的实践为记录信息的信息结构显性做出了重要的贡献,说明了信息结构显性的可能性和重要性,但还不是直接的信息结构显性。其差别主要在于出发点不同,沃森是针对 Jeopardy!这个特定的节目所可能用到的信息,并与特定场景推理功能结合而展开的信息结构显性化。本节所讨论的信息结构显性是针对一个信息集合,或者说针对一个客观存在,将其内在的(隐性的)含义和逻辑关系显性化,可能这一结构化过程的某些阶段或环节是针对特定应用的,但即使是针对特定的目的,其产生的结果是可以同其他围绕这一特定信息集合的信息结构显性无缝整合在一起,即按照统一的规范实现。



图 3.20 沃森与 Ken Jennings 和 Brad Rutter 在 Jeopardy!中的对决现场情景⁶

⁶ 参见【果壳网:寻找 Watson】<http://discover.news.163.com/special/watson/>及
http://www.guokr.com/article/5152/?_block=article_interested&_pos=0&rkey=0f14.

经过上述对人类数万年来对信息结构显性努力的历史性概述，我们可以得出信息结构可显性和信息结构显性的动力客观存在这两个重要的结论。

定理 3.1: 信息结构可显性定理。任何记录态信息集合所包含的全部含义及其逻辑关系是可以显性表达的，信息结构显性的程度能够与智能体的认知水平一致。为证明这一定理，把既有记录态信息集合分成两组，一组是人类的创造物，即是由人类认知结果转换的记录态信息，二是没有经过人类认知过程直接形成的记录态信息。对于前者，形成的记录本身就是部分显性结构的，对于其中没有按显性结构的规则表示出来的部分，必然存在一组人，至少作者本人是可以将其形成显性结构的；对于后者，任何直接记录都是有目的的，都会按照记录的目的给予显性化，有些记录下来的信息只有少数人可以通过工具理解，有的记录只能部分理解，但只要是理解的，就是可以第一类记录态信息，或转换为显性信息结构。这两者的转换，需要一个强条件，一个弱条件。强条件是必须存在结构显性的需求，没有需求，一般动力就不能转化为具体的动力，只有具体的动力才会变成信息结构显性的实际行动。弱条件是必须存在结构显性的必要环境，人力、资金、工具、处理系统等。之所以将之称为弱条件是因为只要动力在，这样的条件一定可以形成。

定理 3.2: 信息结构显性的动力是客观存在的。生物智能体应对生存压力和探索未知的渴望、对已有信息有效利用的需求，构成了促进信息结构显性的动力链环，推动自在态信息转换为自有态信息、自在态和自有态信息转换为记录态信息，记录态信息本身在一定程度上是结构化的，对获取信息和利用信息的动力又转换为持续追求信息结构显性化的动力。

推论 3.19: 信息结构显性是一个长期的发展过程。信息结构显性对认知能力的依赖、对外部约束环境的敏感构成了信息结构显性的约束，而可显性的本质和客观存在的动力又会推动其持续完善，因此，信息结构显性必然是一个长期发展，不断进化和完善的过程。

3.8 显性信息结构的完备性

信息结构显性是一个长期的演进过程，演进的方向是逐步走向完备，什么是完备，完备性的测度、参照系和跨信息结构的运动是本节的内容。

3.8.1 信息结构完备的相对性

3.8.1.1 信息结构的绝对完备和相对完备

定义 3.30: 信息结构的完备性。若称 A^+ 为信息集合 A 的完备显性结构，则 A 中所有的含义、功能、连接都被显性、科学地表述出来。

显然这是一个理想的绝对完备性定义。追求显性信息结构的完备如同人类追求对物理世界宏观和微观认识的极致，在很长的历史阶段内，都是可望而不可即的。无论从认识论的角度还是从实际使用需求的角度，追求信息结构相对完备，或者说在追求相对完备的过程中实现绝对完备是更为重要和现实的路径。

相对完备性有多个维度，取决于使用什么参照系。相对于客观存在的大小、规模，是范围相对性；相对于对应信息集合内含的语义，是逻辑关系相对性；相对于事件发生的场景，是环境相对性；相对于特定目的，是任务相对性；相对于处理能力，是功能相对性。

3.8.1.2 相对于客观存在，范围相对性

任何信息集合对应一个客观存在，但除开自在态信息，自有态和记录态的信息一般不能完整表述对应的客观实在，或者说相对于客观存在，信息集合一般地也是不完整的。

前面已经明确，信息结构完备性的参照系是客观存在。客观存在间尽管存在不同的联系，但从认识或利用的角度看，总是区分出一定的范围。所以，

信息结构也同样是这样一个给定范围，边界确定的客观存在的表述，而不是泛指。

如同本章第四节的几个例子所示，客观存在可以是一个特定的道路状态，也可以是同一段道路一段时间的状态，还可以是若干条相邻道路的状态；可以是视网膜，眼睛，整个视觉神经系统，还可以是整个感觉神经系统。相对于客观存在，任何一个特定维度的范畴都是局部。

3.8.1.3 相对于信息集合，逻辑关系相对性

一个信息集合和与其对应的客观存在之间不仅具有映射不完整，更重要的区别是信息集合的本质是逻辑存在，不是实体存在，信息集合的实体只是载体，而不是其本意。因此相对于信息集合，信息结构的相对完备性是指在信息集合所规定的范畴内，逻辑关系完备程度的相对性。一个信息集合，其中存在的含义、相互间及与外部连接的逻辑关系是确定的，对应的显性信息结构所表述出来的也是确定的，但一般并不完备，这里不仅有认知水平的原因，还有后面论及的场景、目的等因素，更有成本、时间、人力等多种原因。

在影响逻辑关系相对性诸多要素中，认知水平是关键因素。如果认知水平达不到，即使有需求、有必要，而且时间、成本、人力都可支持，显性信息结构的完备度最高只能达到认知水平的顶点。例如，表 3.3 中关于神经系统视觉功能信息单元组件（部分），完全依赖于学术界对视觉功能研究的成果。如果从视觉细胞感知的形成到视觉信息传递过程、视觉信息记忆以及知觉合成的信号传递和汇聚能够全面把握，就可能生产媲美美人的机器视觉。

3.8.1.4 相对于特定场景，环境相对性

特定场景是指一个信息集合围绕一个场景形成或一个任务所处的场景，一般说来，这个场景是动态的，而不是静态的。这也是相对于特定场景的显性信息结构完备性与相对于客观存在的完备性的区别所在。

许多信息集合围绕特定场景形成。如一次会议、一场演出的录音、录像或速记，一个人在山区乡村夜晚散步时的感知和调节，玉米授粉过程碰上雷阵雨，等等。

场景是信息生成最主要的类型，区分信息结构显性的环境相对性，是许

多应用的需要。环境相对性与任务相对性有联系，也有区别。如果是指任务在特定场景中发生，而这个场景又与围绕特定场景形成的信息集合一致，则恰好重合；但很多场合又不尽一致，除非是围绕任务而设定的场景。

从信息结构显性看，即使是重合的场景，信息集合与任务信息集合，最终产生的结构并不一定一致。这是显性的目的不同，对于场景信息集合的结构显性，重在逻辑关系的完备性；对于任务信息集合的结构显性，重在与任务相关的含义和功能的完备性，场景中与任务无关的逻辑关系，则一般不做显性表述。例如，在表 3.5 色素上皮细胞基础部分结构描述中，这一信息结构显性的目的是描述视觉神经系统的过程，没有在细胞级做进一步展开，因此在描述部分只阐述了防止光线的反射，消除来自巩膜的散射光线，避免光感受器受到过度光刺激，为光感受器细胞提供营养，维持光感受器细胞视色素的正常代谢等过程性功能描述，而没有就细胞功能实现的生物和化学细节展开。

3.8.1.5 相对于特定目的，任务相对性

相对于特定目的是指信息结构显性既不是源于一个具体的客观存在、一个信息集合、一个特定场景，而是源于具有明确目的的任务。

上节提到的计算机系统“沃森（Watson）”在 2011 年 2 月 14～16 日美国“Jeopardy”电视娱乐节目中战胜两个该节目胜率最高的对手，就是以“Jeopardy”电视娱乐节目中涉及的信息为对象，以节目实际进行的过程为场景，既对 2 亿页、数百万份资料按照节目问答的特殊要求进行了结构显性，还对节目进行环境中的语音识别、流程规则等功能同样做了结构显性，还对搜索、匹配、问答风格设定等算法的结构显性化，形成了历史上可能最复杂的针对任务的信息结构显性。尽管这一信息结构显性十分复杂，已经到了超越人类强者的水平，但相对于任务，还是不完备的。尽管赢得了胜利，但存在很多需要进一步完善的地方，从语音识别、信息覆盖到算法、问答风格，都不完备。

相对于任务的结构显性，一般地并不要求完备，也很难达到完备。就上面的例子，就是将信息量再增加 10 倍，还是不能保证没有失误，不能保证对手不得分。这是特殊例子，就一般而言，针对任务的信息结构显性通常缺乏

足够的时间、人力和资金支持，从而追求在给定时间、经费和人力的前提下达到最好的结果，也就是不追求理想的绝对完备，而是追求基本满足任务需求的相对完备。

3.8.1.6 相对于处理能力，功能相对性

信息结构是客观存在的逻辑抽象，如同本章 3.4 节中几个示例所展示的，不仅要显性描述各实体的属性，还要描述该客体拥有的功能及功能的实现。从这个角度看，信息结构也是一个仿真模型，一个完备的功能描述，可以在仿真系统中重复该客体的全部功能。

功能相对性是指显性描述出来的功能及功能实现过程与信息集合或客观存在内含的功能及其实现过程不完整或不一致。不同的客观存在或信息集合所含功能存在巨大的差异。对于功能比较简单或人类认知研究比较透彻的客观存在，其间的差异性小一些；反之则大一些，而研究尚未揭示的功能和功能实现的过程，当然不能显性表述出来。

功能相对性还取决于任务的需求。同逻辑关系一样，一项特定的任务并不一定要用到对应客观存在的所有功能或功能的全部过程，本章 3.4 节中的示例分析，都可以得出这样的结论。

归结起来，显性信息结构的理想绝对完备是给出了一种比较的基准和方向，而实际存在的显性信息结构是不完备的，但这种不完备的显性结构，已经能够基本满足很多实际应用的需求。显性信息结构不完备的根本原因是人类认知水平没有达到全部揭示的程度；而在大部分特定目的的任务和场景中，任务本身只要求相对完备，加上时间、人力、资本等因素，决定了显性信息结构的相对完备。

3.8.2 信息结构完备的测度

显性信息结构的完备既是一个长期的发展过程，又因为具体环境和目的得不同而存在不同的要求和约束，需要有一个测度来对已经存在的显性信息结构进行完备性度量。显性信息结构完备的测度一个系列，分别测度信息结

构的不同维度，这些测度包括：颗粒度、详尽度、符合度、重现度和进化度。

显性信息结构是抽象的，是智能活动的产物；显性信息结构又是动态的，因为它的对应物客观存在是变化的，自身的显性是一个完备的过程。测度既对应于理想的完备结构，为结构显性的演进指出方向，也对应于相对完备结构，对实际存在的信息显性结构或工作进行评价。

3.8.2.1 颗粒度

颗粒度衡量的的是一个具体信息结构单元划分的大小，及达到理想的完备结构最小可区分单元的距离。

定义3.31: 设被测度显性信息结构为 A_l ，其对应信息集合的完备结构层次深度为 nl ， A 的层次深度为 ml ，则称 A 的颗粒度为 ml/nl 。如 nl 为 15， ml 为 6，则 A_l 的颗粒度为 0.4，达到理想完备结构的距离是 9。因此，颗粒度的取值范围是 $0 \sim 1$ ，一般不为 0，1 则是完备结构。距离和颗粒度的数值成反比关系。

推论 3.20: 相对颗粒度。由于相对于一个具有一定复杂程度的具体信息结构的理想完备结构绝大部分还不存在，颗粒度的测度需要一个可操作的办法，本文称之为相对颗粒度。设被测度显性信息结构为 A_l ，其对应信息集合的评介完备结构层次深度为 nl' ， A_l 的层次深度为 ml' ，则称 A_l 的相对颗粒度为 ml'/nl' 。如 nl' 为 8， ml' 为 6，则 A_l 的相对颗粒度为 0.75，达到评介完备结构的距离是 2。其中作为基准的评介完备结构是对相应信息集合在已有显性信息结构的基础上由专家评估出的全部层次深度。

前面已经讨论，完备度的相对性有五个不同维度，分别是范围、逻辑关系、环境、任务和功能，这里分别记范围被测度显性信息结构为 A_{l1} ，其对应信息集合的评介完备结构层次深度为 nl_1 ， A_{l1} 的层次深度为 ml_1 ，则称 A_{l1} 的范围相对颗粒度为 ml_1/nl_1 ；记逻辑关系被测度显性信息结构为 A_{l2} ，其对应信息集合的评介完备结构层次深度为 nl_2 ， A_{l2} 的层次深度为 ml_2 ，则称 A_{l2} 的逻辑关系相对颗粒度为 ml_2/nl_2 ；记环境被测度显性信息结构为 A_{l3} ，其对应信息集合的评介完备结构层次深度为 nl_3 ， A_{l3} 的层次深度为 ml_3 ，则称 A_{l3} 的环境相对颗粒度为 ml_3/nl_3 ；记任务被

测度显性信息结构为 A_{14} ‘，其对应信息集合的评介完备结构层次深度为 nl_4 ‘， A_{14} ‘的层次深度为 ml_4 ‘，则称 A_{14} ‘的任务相对颗粒度为 ml_4 ‘/ nl_4 ‘；记功能被测度显性信息结构为 A_{15} ‘，其对应信息集合的评介完备结构层次深度为 nl_5 ‘， A_{15} ‘的层次深度为 ml_5 ‘，则称 A_{15} ‘的范围相对颗粒度为 ml_5 ‘/ nl_5 ‘。

3.8.2.2 详尽度

详尽度衡量的的是一个信息集合内含的属性和功能单元被揭示的覆盖度以及达到详尽的距离，也就是度量可能存在的属性和功能单元有多少被揭示了出来，有多少还没有被揭示出来。

定义 3.32: 设被测度显性信息结构为 B ，其对应信息集合的完备结构各级可区分单元数为 nd ， B 已区分出的单元数为 md ，则称 B 的详尽度为 md/nd 。如 nd 为 15000， md 为 6000，则 B 的详尽度为 0.4，达到理想完备结构的距离是 9000。同样，详尽度的值与距离成反比。

推论 3.21: 相对详尽度。设被测度显性信息结构为 Bd ‘，其对应信息集合的评介完备结构层次深度为 nd ‘， Bd ‘的层次深度为 md ‘，则称 Bd ‘的相对详尽度为 md ‘/ nd ‘。如 nd ‘为 9000， m ‘为 5400，则 bd ‘的相对详尽度为 0.6，达到评介完备结构的距离是 3600。其中作为基准的评介完备结构是对相应信息集合在已有显性信息结构的基础上由专家评估出的全部应该具有的属性和功能单元。

如同相对颗粒度，相对详尽度也要从范围、逻辑关系、环境、任务和功能五个维度分别定义。记范围被测度显性信息结构为 Bd_1 ‘，其对应信息集合的评介完备结构单元总量为 nd_1 ‘， Bd_1 ‘表述的单元数为 md_1 ‘，则称 Bd_1 ‘的范围相对详尽度为 md_1 ‘/ nd_1 ‘；记逻辑关系被测度显性信息结构为 Bd_2 ‘，其对应信息集合的评介完备结构单元总量为 nd_2 ‘， Bd_2 ‘表述的单元数为 md_2 ‘，则称 Bd_2 ‘的逻辑关系相对详尽度为 md_2 ‘/ nd_2 ‘；记环境被测度显性信息结构为 Bd_3 ‘，其对应信息集合的评介完备结构单元总数为 nd_3 ‘， Bd_3 ‘表述的单元数为 md_3 ‘，则称 Bd_3 ‘的环境相对详尽度为 md_3 ‘/ nd_3 ‘；记任务被测度显性信息结构为 Bd_4 ‘，其对应信息集合的评介完备结构单元总数为 nd_4 ‘， Bd_4 ‘表述

的单元数为 $md4'$ ，则称 $Bd4'$ 的任务相对详尽度为 $md4'/nd4'$ ；记功能被测度显性信息结构为 $Bd5'$ ，其对应信息集合的评介完备结构单元总数为 $nd5'$ ， $Bd5'$ 表述的单元数为 $md5'$ ，则称 $Bd5'$ 的范围相对详尽度为 $md5'/nd5'$ 。

3.8.2.3 符合度

符合度是衡量一个显性信息结构所揭示的对应信息集合或客观存在内含的属性和功能的一致性，或者说，在显性结构中表述出来的属性和功能单元与实际存在是相同的还是有差异的。

定义 3.33: 设被测度显性信息结构为 E ，其对应信息集合完备结构的单元总数为 nc ， E 表述的单元数为 mc ，其中，经评价证实一致的单元数为 mcc 则称 E 的符合度为 mcc/nc 。如对应信息集合的完备结构各级可区分单元总数为 10000， E 已区分出的单元数为 6000，经评价一致的单元数为 5000，则 E 的符合度为 0.5。

推论 3.22: 相对符合度。设被测度显性信息结构为 E' ，其对应范围、逻辑关系、环境、任务和功能五类信息集合的评介完备结构单元总数分别为 $nc1'$ 、 $nc2'$ 、 $nc3'$ 、 $nc4'$ 、 $nc5'$ ，评价一致的单元数分别为 $mc1'$ 、 $mc2'$ 、 $mc3'$ 、 $mc4'$ 、 $mc5'$ ，则范围、逻辑关系、环境、任务和功能五类信息集合的相对符合度分别为 $mc1'/nc1'$ 、 $mc2'/nc2'$ 、 $mc3'/nc3'$ 、 $mc4'/nc4'$ 、 $mc5'/nc5'$ 。

符合度应由三个部分构成，一是属性描述符合度，主要体现在 3.8.1.1 和 3.8.1.2；二是功能描述符合度，功能单元的符合度在本节讨论，功能过程的一致性讨论见 3.8.1.4，三是进化结构与描述符合度，在 3.8.1.5 中讨论。

3.8.2.4 重现度

重现度是衡量一个信息集合内含的功能单元及功能实现的过程（也以功能单元的形式表述）被揭示的正确性。衡量如果将揭示的功能及过程有仿真模式运转，与客观存在实际进程的一致性。

重现度有三个组成部分，一是功能单元描述的正确性，属于符合度的一个部分；二是过程的一致性，即一个功能在实现过程按信息结构描述的顺序和规范产生的结果与实际客观存在的进程和结果的一致性；三是在一个动态的发展过程中，信息结构显性过程与客观存在发展过程中功能的一致性，能

重现动态进程及其结果，这个部分在 3.8.1.5 中讨论。

对于功能实现过程的重现度，取决于显性结构中的三个要素，一是单元中功能的实现过程，二是信息结构中整体功能的实现过程，三是整体功能实现的规则。根据这样的分析，我们将重现度定义如下。

定义 3.34: 设被测度显性信息结构为 F ，其对应信息集合完备结构的功能进程有 nr 条， F 表述的功能进程数为 mr 条，其中，经评价证实结果正确的条数为 mrr ，则称 F 的重现度为 mrr/nr 。如对应信息集合的完备结构各信息单元及信息结构各层次的功能进程数为 5000， F 已区分出的功能进程数为 4000，经评价一致并得到结果验证的单元数为 3000，则 F 的符合度为 0.6。

推论 3.23: 相对重现度。设被测度显性信息结构为 F' ，其对应感知、连接、处理、进化四类功能的评介完备结构功能进程数分别为 $nr1'$ 、 $nr2'$ 、 $nr3'$ 、 $nr4'$ ，评价一致并得到结果验证的功能进程数分别为 $mr1'$ 、 $mr2'$ 、 $mr3'$ 、 $mr4'$ ，则范围、逻辑关系、环境、任务和功能五类信息集合的相对重现度分别为 $mr1'/nr1'$ 、 $mr2'/nr2'$ 、 $mr3'/nr3'$ 、 $mr4'/nr4'$ 。

3.8.2.5 进化度

进化度衡量的是一个信息集合的信息结构是否具备良好的规则，保证信息结构的发展能朝着理想绝对完备方向并能与其表述的客观存在的发展保持相同的方向，换言之，进化度就是衡量方向的一致性。

保证方向一致性的工具就是确立好的信息结构发展规则。发展规则的好坏需要发展过程的结果来评价。发展规则是指一个显性信息结构在一开始设定的描述和功能演进的规则，一般应包括以下四个内容：一是单元结构和整体结构的包容性和一致性；二是感知、连接、处理三大功能的包容性和一致性；三是整体进化规则的预见性和包容性；四是跨信息结构连接和协同规则的包容性和预见性。

一个显性信息结构的完善是一个过程，一个信息域或一个文明的显性信息结构完备更是一个漫长的历史过程，进化的规则在其中扮演关键的角色，在下一章将更加深入地讨论这个主题。

3.8.3 信息结构间的关系及完备性

客观存在的相互关联，可以根据实际任务或分析场景要求，划出一个独立的客观存在，但与周边的关系依然存在。一个信息结构是对一个已经区分出来的客观存在的描述，但相关客观存在对应的信息结构间的联系依然存在。在一个信息结构完善的过程中，会利用、参照、整合其他信息结构的成果，也可能切分出一个子集成为独立的信息结构。需要专门讨论信息结构之间的关系及在信息结构完备中的作用。

3.8.3.1 信息结构的体系

前面我们讨论了信息单元和作为个体的信息结构，接下来，将进一步讨论作为整体的信息结构，也就是显性信息结构的体系。

显性信息结构反映的是一个因不同原因而切分出来的客观存在的不同层次的含义及其相互间逻辑关系，而这个客观存在与其他客观存在具有许多联系。显性信息结构既要如实反映这些联系，还要通过必要的功能和规则来实现信息结构体系间的联系和协同发展，信息结构体系就是在一个给定的范畴内信息结构间要确定的关于连接和协同发展的功能和规则，使一个个独立发展、各自具有特定功能的信息结构形成一个与客观存在联系和信息结构演进规律一致的体系。

显性信息结构的体系是一个包含信息结构间联系及其功能实现的层次性架构，图 3.21 是其示意图。

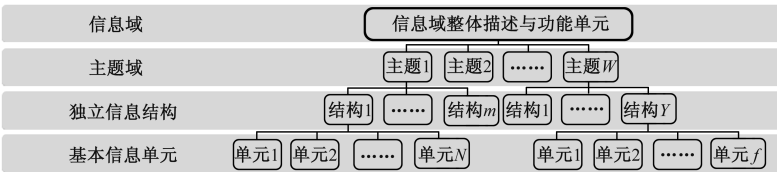


图 3.21 一个信息域的信息结构体系示意图

如图 3.21 所示，在同一信息域内，至少可以分成四层，即信息域、主题域、独立信息结构层、基本信息单元层。其中，信息域一般是指使用相同外

壳的记录信息集合和主体的综合体，还包括物理隔离或权力空间控制而构成的记录信息不共享的信息域；主题域是指同一信息域中依据不同语义类和任务类需求形成的相对独立的信息子空间；独立信息结构是指每一个主题域下形成的具有独立功能的显性信息结构；基本信息单元就是构成独立信息结构的基础构件。

从记录信息显性结构的全局看，还有一个半层次，一个层次是信息结构全域，半个层次是全域中的可转换信息域。所谓全域是指所有记录信息的信息结构，不管是完全显性的还是局部显性的，这里所指全域，是指地球文明，不包括地外文明。所谓可转换信息域是指存在一定的主体或客体，他们能够使不同外壳或绝对分割的物理空间、权力空间内的记录信息构成的信息域之间可以进行一定程度的转换，这样的信息域间不是各自封闭的独立信息域，二是存在相互渗透、协同和利用的半封闭信息域。可转换信息域相对于不可转换信息域，其中的信息域都是各自封闭的。在人类文明发展的早、中期，存在大量不可转换信息域，现在也还有少量不可转换信息域，至于地外文明，应该属于不可转换信息域。

一个信息域内信息结构的体系化，需要实现信息结构间的可互用。互用可分为两种类型，一种是满足一个信息结构在完成一个任务时，可利用其他信息结构的成果；另一种是信息域整体信息结构完善过程中的互用和一个信息结构完善过程中的互用，加快信息结构的组合、分解、进化。

信息结构体系通过如图3.14中整体描述与功能单元实现这样的目的或功能。信息域与主题域都有一个具有全局功能的整体描述和功能单元，这里的单元不一定是一个，视实际需求，可以是一组，并保留一种弹性的结构，可以在进化过程中完善增减。整体描述主要用于协同，特别是跨主题域、跨信息域的协同，而功能主要是用于本域内信息结构间的感知、连接、处理功能和规则的一致性及其实现，信息结构组合和切分功能的实现等，具体将在以下有关部分讨论。

在一个信息域内信息结构间的互用和协同进化，要区分两种场景，一是同一主题域，二是跨主题域。同一主题域的实现通过本主题域的顶层描述和功能单元来实现，跨主题域则需要本主题域的顶层描述与信息域顶层描述、

相关主题域顶层描述的协同才能实现，因此，在一个信息域必须对整体的描述和功能的规则和格式有严格的规定，这样才能实现期待的功能。这些规则主要包括命名、地址的规范，感知、连接、处理功能局部与全局一致，并各有侧重，各信息结构自身增长完善与全局和相关信息结构的生长和完善的协同实现准则及处理功能安排，各信息结构在生长和完善过程中相同对象的处理出现冲突的处置原则及其实现，等等。

跨信息域信息结构体系需要实现的功能与一个信息域内的基本相同，主要差别是如何把外壳不同的转换作为一种演进功能引入整体功能中。

3.8.3.2 组合与分解

3.3.4.2 讨论过信息单元的组合，这里进一步讨论更广泛的组合与分解。

组合与分解是信息结构间和信息结构内最频繁发生的事件，本节只分析同一信息域内发生的组合与分解。

定义 3.35：信息单元的组合。信息单元的组合是指两个或两个以上信息单元的全部或部分合并为一个新的信息单元。

定义 3.36：信息单元的分解。信息单元的分解是组合的逆运动，是指一个信息单元或几个信息单元分解出一个或几个新的信息单元。

定义 3.37：信息结构的组合。信息结构的组合是指两个或两个以上信息结构的全部或部分合并为一个新的信息结构。

定义 3.38：信息结构的分解。信息结构的分解是指从一个信息结构或几个信息结构分解出一个或几个新的信息结构。

分解和组合的边界并不总是清晰，有时候会产生重合的现象。如信息单元组合的特例，若个个信息单元组合成一个新的信息结构，从原来的信息结构中分离出来，这就符合信息结构分解的定义。

信息单元和信息结构组合与分解的动力有两个来源，一是信息结构完备过程中基于语义或功能发展变化而产生；二是基于完成信息结构面临的任务而组合或分解。

不管源自什么，组合或分解产生的新的信息单元或信息结构，有的会长期稳定存在，有的则是临时的。长期稳定存在，说明新单元或新结构是有用

的、成熟的。首先是有用的，新的单元或结构在表述客观存在与/或完成任务是具有不可替代的作用；其次是成熟的，即表述的属性或功能与客观存在有较高的相似度，至少比已经存在的显性结构中表述得要更加正确。之所以是临时存在，有的是因为任务结束了，而又不是在语义上不可或缺的单元或结构；有的是因为成熟度不够高，进过一段时间的检验，证实可以有新的单元或结构替代，或正确性低到不能接纳到信息结构中。

3.8.3.3 信息结构差

由于某种任务将大脑中的认知信息或记录信息形成显性结构存在，这是显性信息结构的起点，而发现结构差、明确结构差，减少结构差，则是信息结构走向完备的起点，信息结构差在信息结构完备，也是信息增长的一个关键环节。

结构差是衡量两个信息结构之间的不同，而信息结构之间的差异由四个部分构成：相同的、不一致的、多的和少的。据此，我们定义信息结构差如下。

定义 3.39: 假设有两个信息结构 A 和 B，经过比较分析，信息结构 A 共有信息点 N_a 个，信息单元 M_a 个，信息结构 B 共有 N_b 个信息点， M_b 个信息单元；两个信息结构间相同的信息点为 n_1 个，相同的信息单元数为 m_1 个；不一致的信息点为 n_2 个，信息单元为 m_2 个；A 比 B 多 f_1 个信息点， f_2 个信息单元，而 B 中又有 h_1 个信息点和 h_2 个信息单元是 A 中所没有的。则 A 相对于 B 的信息结构差 AdB 为：

$$AdB = Nad + Mad = (n_2 + h_1) / (N_a - n_1 - f_1) + (m_2 + h_2) / (M_a - m_1 - f_2)$$

同理，B 相对于 A 的信息结构差 BdA 为：

$$BdA = Nbd + Mbd = (n_2 + f_1) / (N_b - n_1 - h_1) + (m_2 + f_2) / (M_b - m_1 - h_2)$$

式中， Nad 和 Mad 分别表示 A 相对于 B 的信息单元差和信息结构差， Nbd 和 Mbd 则是 B 相对于 A 的信息单元差和信息结构差。 Nad 、 Mad 、 Nbd 、 Mbd 均大于 0，小于 1，数值越大，结构差越大。

上面两个表达式还不足于精确度量两个信息结构之间的结构差。正如前一节讨论信息结构冗余的时候指出的，信息单元不是信息结构的基本单位，基本单位应该是信息点，而且信息点也有不同的性质，具有不同的功能，一

个信息结构，一个主题域和一个信息域，都应有综合性的单元来描述整体，规范进程，这些对精确讨论结构差都有重大的影响。

需要引入新的计量单位，使上面表达式真正可计算。可计算的基础是信息点，用义元（mit）来表示。用义元，一来表示是计量信息含义的最小单位，二来与 bit，信息处理最小的物理计量单位之间建立一种形式的关联。

在 3.3 节中，已经规范一个信息单元有四类不同的组成部分：基础、感知、连接、处理，每一类又由若干个信息点构成。对这四类信息点，分别用 mitb、mits、mitc、mitp 来表示，则有：

$$\begin{aligned}\sum \text{mitb}_i &= \text{MITb}, \quad \sum \text{mits}_i = \text{MITs}, \quad \sum \text{mitc}_i = \text{MITc}, \quad \sum \text{mitp}_i = \text{MITp}, \\ \text{MITb} + \text{MITs} + \text{MITc} + \text{MITp} &= \text{UMIT}.\end{aligned}$$

UMIT 是一个信息单元的信息点的总和。如果需要更加清晰地区分不同构成类型信息点在信息单元中的作用，可对四类信息点，甚至同类信息点中的特定点给予加权，用加权后的计量来满足不同的测度目的。

3.9 关于信息结构的进一步讨论

信息结构是全书的核心概念。信息结构是一个中介，将自在态、认知态、记录态信息连接在一起。信息结构是智能活动的结晶，也是加快、引导智能活动的标杆。信息结构是智能活动的产物，但具有不同于任何物体的特征并能不依赖智能生物体发展。本节对这些结论作简要的分析。

3.9.1 信息结构是连接不同态信息的纽带

在无穷无尽的物质运动过程中，客观存在的属性和运动通过各种物理介质显示其含义和逻辑关系；遗传基因在漫长的进化过程中积累信息，并把生命特征和生长过程通过信息精确地延伸和控制；生物体通过获取外界信息并做出反应，并在长期的演进中形成了神经系统、大脑等信息处理的功能和器官；这些功能和器官在外部环境的刺激下，逐渐形成语言、文字和概念，使

得信息有了可以独立于有生命和无生命的客观存在，记录信息是客观实在的反应，是认识“信息”这个词的载体。

我们以光合作用这个客观存在为例，说明信息结构如何成为认识信息，连接不同态信息的纽带。图 3.22 由 a、b、c、d、e 五个部分组成。其中 a 是加上标识的一个光合作用过程的客观存在的图片，如果去掉标识，就是一个普通的叶片，光合作用就在这个叶片中发生。b 是光合作用如何实现的一个框架模型，显示了光透过气孔，进入叶绿体，通过基质类囊体的光合膜把光能转化为生物能。c 是关于光合作用的一段文字描述。d 在蛋白质层次展示了光能如何转化成生物能。e 则进一步细化了光合作用蛋白复合体的 3 维结构，在分子级刻画了其晶体结构。

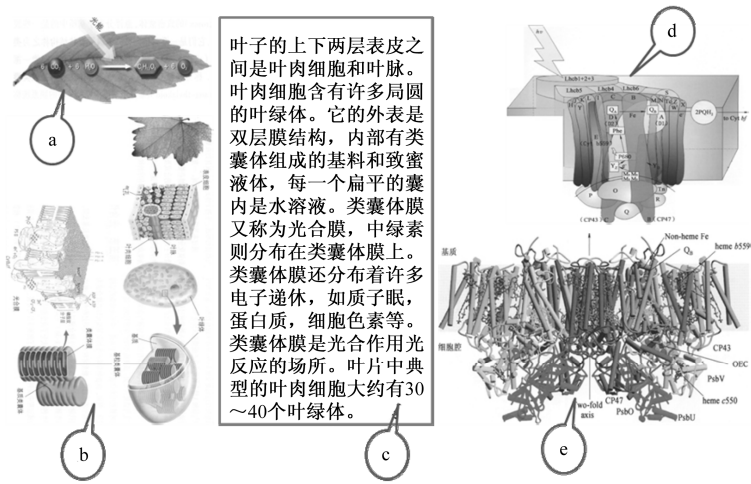


图 3.22 光合作用不同形态和不同表示方式的信息示意图⁷

显然，a 是自在态信息，也是部分生物体自有态认知信息的一个部分，b、

7 图中，a 和 b 源自 基础生命科学（第二版），吴庆余编著，高等教育出版社，2006 年，第 113 和 114 页 c 根据基础生命科学（第二版），吴庆余编著，高等教育出版社，2006 年，和植物生理与分子生物学（第四版），陈晓亚、薛红卫主编，高等教育出版社，2012 年有关内容编写 d 和 e 分别源自植物生理与分子生物学（第四版），陈晓亚、薛红卫主编，高等教育出版社，2012 年，第 237 和 238 页。

c、d、e 均是记录态共有信息，又是认知态信息的一种存在模式（学者的记忆）。这些不同态的信息具有相同的含义、功能和含义、功能间的逻辑结构，指向相同的内容但没有客观的连接纽带，只能通过学者的大脑来辨析，又回到了认知态信息，依赖主体的认知能力。

从信息结构来看，a 去掉标记部分，是纯粹的隐性结构，b、c、d、e 总体上是隐性结构，但具有一定显性信息结构的基础，尤其是 c；如果围绕光合构建作为一个显性信息结构，将从不同类别的分子、蛋白质到承担不同功能组织，从光线的进入到光能转化为生物能的全过程，如图 3.23 所示，信息结构显性化，则可以得到一个包括数千个信息单元，以如同本章 3.4 节中的例示，以信息结构总表为基础，逐次展开不同类型和层次的信息单元，文字描述、实体图形、模型及注解并重，把迄今为止的相关认知成果结构化，就形成了一个连接自在态、自有态、记录态信息及相应隐性信息结构和信息集合的纽带，而且也理解信息提供了真实的基础。

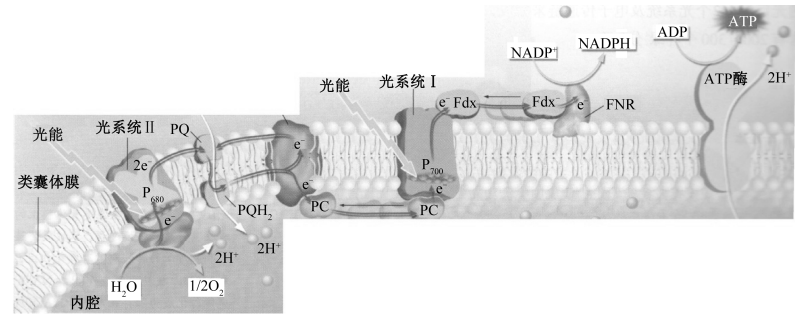


图 3.23 光能转化示例图⁸

3.9.2 信息结构对加快和引导科学发展具有重要作用

信息结构不仅为不同信息存在形态提供了连接纽带，更为文明和智能的发展提供了最重要的载体和方向。假定已经存在一个关于光合作用的最简完

⁸ 基础生命科学（第二版），吴庆余编著，高等教育出版社，2006 年，第 118 页。

备显性信息结构，这里，完备相对于当前关于光合作用的最高认知水平，那么无论是教学还是农业工程，是论文发表审核还是科研成果鉴定，都有了一个基准，准确的还是有偏差的，有新的发现还是已有结论的翻新。

显性信息结构还有一个重要的功能，就是从两个方面加快科研创新的进程。一是对照这样的结构比查找浏览文献更快、更可靠，这是显然的；二是它的共有性，可以使更多的人能够借助他人的研究成果探索前进，科学的进步与参与的人数成正比，尤其是与能站到科学前沿的学者的数量成正比，显性完备最简信息结构就能最大限度地使一切有可能的人最快走到科学前沿。

3.9.3 信息结构的一些重要特征

信息结构是智能，或者说是文明进化的结果，是一个高度抽象又具有充分的实感，它极其重要，但迄今为止还没有一个完备显性信息结构产生。这是由信息结构本身的特征所决定的。

信息结构有许多其他精神或物质存在所没有的特征，这里罗列几个最重要的。一是它本身是以记录态存在的信息，但它包含了处理功能，这个功能不仅使它能指导重现其所表述的客观存在的功能，还使得符合大规模、不同性质的多任务信息处理规律的多中心处理成为可能。一个信息结构既要完成自身完善的学习型处理任务，又要完成外部对其提出的各类任务，处理的要求可能是全局的，既与整个信息结构相关，也可能是局部的，只与少数信息单元相关，多中心处理成为信息结构的必然要求，也成为其重要特征。二是大数之维。在 3.4 节中举了街景这个例子，十分简单，但透过这个常见、极其简单的例子，可以看到一个信息结构具有大量的维度。要穷尽街景中的维度，肯定超过百，达到千这个数量级也是可能的。同一个事物，置身于不同的场景或任务，就是一个不同的维度，不同场景使同一信息的描述和功能都会产生不同理解或含义。三是不定之值。多维度、多场景、成熟度不一等原因，决定了信息单元中很多信息点都是不定值。从属性的描述到相应的物理地址，从连接的对象到连接的条件，从处理的目的到处理的模式，存在大量需要动态和随机赋值，因此，在信息结构中，大量的信息点利用函数进行表述，甚

至还有不同或不定量纲，函数也可能是指代的。四是多结构并存。显性信息结构产生的原因不同，处理的任务或应对的场景不同，颗粒度和详尽度，重现度和冗余度等要求也不相同，因此，对应相同的客观存在，一定会有多个，甚至很多个信息结构的存在。五是繁杂的连接和连接的多样性。不管是最简结构还是冗余结构，由于客观存在联系的广泛性和任务、场景的多样性，一个信息点、一个信息单元以及其上的各个层次，多存在大量的联系，一个点（广义的点）可能存在10的3次方、4次方、5次方甚至更多的连接。六是交互发展。信息结构走向完备是必然的，走向完备的路径是多样的，但在信息结构完备的不同阶段多会使用交互的模式，或者说相互学习的模式。交互的对象也是多元的，可能是人、也可能是系统，还可能使任务中的其他对象。

信息结构的生长是信息增长的主要内容，在下一章将进一步讨论本节的一些内容。

第 4 章

信息的增长

信息以不同载体、不同外壳、不同形态、不同主体、跨越空间和时间，生成、移动、计算、积累、利用，历经数十亿年不断进化，从自在态信息走向记录态信息，从隐性结构走向显性结构，并最终走向独立信息空间。本章将讨论这一增长过程。

4.1 信息增长的定义和发展阶段

4.1.1 信息增长的定义

自在态信息在持续增长。从大爆炸以来 100 多亿年的时间里，宇宙起点的信息还可能观察到，物质运动又持续产生着新的自在态信息，无限的自在态信息沿着时间轴持续增长。

自在态信息在持续增长。具体物种的遗传信息在变异中发展，而没有变异的依然存在于不同个体的遗传信息中，甚至存在于变异个体的自身遗传信息中；认知信息因个体的增加而增长，更因认识水平的提升而增长。

记录态信息在持续增长，从语言和文字的诞生到记录技术的发展，从少数人的舞台到全民参与，从模拟信号到数字，从印刷品到电子存储，人们惊呼，进入信息爆炸时代。这个名词，应该只适用于记录态信息，相对于自在态信息，记录态信息够不上沧海一粟。

信息在持续增长，但本章讨论的信息增长指什么，需要明确定义。

定义 4.1: 信息增长。信息的增长是指各类信息载体数量的增加和信息结构的完备过程。这一定义覆盖了所有的信息形态和这些信息形态的依托的外壳和载体，这是信息增长量的部分；定义同样包含了信息增长质的部分，那就是信息结构的完备和增长。这里的信息结构包括了隐性结构和显性结构，包括了很不完备和相对完备的，包括了各类信息域和主题域。信息结构的增长既包含已有结构的完备，也包括新增加的信息结构。

对于无限的自在态信息，确实在增长，但在无限中增加一个部分，依然是无限。所以本章不讨论自在态信息自身的增长，只讨论它对自有态、记录态信息的影响和作用。

4.1.2 信息增长的发展阶段

如图 4.1 所示，我们现在观察到的宇宙形成以来，信息的发展经历了五个阶段，正在向第六个阶段演进。这个约束包含两层含义，一是迄今为止可观察到的这个宇宙，二是这个宇宙中迄今为止从地球观察到的文明。

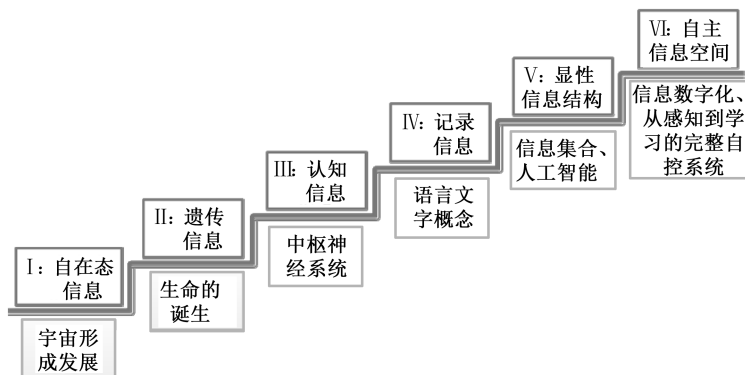


图 4.1 信息发展进化过程

第一个阶段是自在态信息的产生和发展。起点是宇宙大爆炸¹，宇宙诞生以来，伴随宏观和微观的物质运动，自在态信息同时形成和发展。今天的天文望远镜，以宇宙背景微波辐射理论为基础，探索宇宙形成早期的秘密，为这种观察，获取的就是宇宙早期运动的信息。自在态信息不仅在宇宙生成的一瞬间就产生，而且在叠加增长，绵绵不断、无穷无尽。

第二个阶段是遗传信息的产生和发展。如果没有一种过程利用自在态信息，其本身是没有意义的。生命的诞生，生命感知环境信息，对自在态信息实现了两次转化，两次转化推动信息发展上了两个台阶。第一次是生命应对环境的过程中积累并强化了遗传信息的进化，尽管突变是基因变化的主要原因，但有些突变的发生就是因为环境造成，而变异后“适者生存”的选择规

¹ 我们生存的这个宇宙如何形成，目前还没有定论。这里假定 130 多亿年前的大爆炸成为这个宇宙的起点。

则，就主要由环境来承担了。环境是物理状态，但状态与生命作用的过程是信息的过程。遗传信息的产生于发展，无处不在、无时不在的自在态信息具有基础性的作用，没有环境信息无时无处不在的特征，生命的进化没有发生的基础。

第三个阶段是认知信息的形成和发展。自在态信息经由生物体实现的第二次转变是生物体中枢神经系统的进化，大脑的发育和进化，使得高等动物不仅能感知环境信息，并据此做出有利于生存的反应，而且能经过大脑的记忆和学习等功能，形成了完整的认知系统。从生命的诞生到认知系统的基本成熟，经历了 30 多亿年时间。

第四个阶段是记录信息的产生和发展。高等动物认知系统的成熟，在一些动物群体形成了复杂程度不等语言，而人类不仅进化出了语言，更向文字和复杂的概念系统演进这个过程也已经走了超过百万年。在信息发展历史上，语言、文字和系统性概念的产生具有特别重大的意义。从认知的角度看，人类的知识和经验可以透过时空的约束，从个体积累走向群体积累，人类的能力增长跨入新的阶段，分析地球文明发展史，语言、文字和概念体系的形成和发展的加速度作用十分清晰。从信息的角度看，首先是人类认知能力的提升，在质和量两个方面加快了增长的速度，其次是记录信息的形成，信息内在含义有了专门指代的外壳，为信息的自我增长奠定了第一块基石。当然，记录信息量的增长还有技术、经济、社会、文化等其他因素。

第五个阶段是显性结构的形成和发展。文字性记录信息从一开始就具有结构，泥板上的楔形文字，龟骨上的甲骨文，都有语法、句法，甚至布局中的结构，理解语义同样要理解结构。到一篇文献、一本著作，都有一些专门的结构性描述，如目录、关键词等，但这些显性结构都是作者根据自己理解的社会通行规则所做的附着性结构显性，颗粒度很大，范围就在这本书或这篇文献。显性信息结构发展的两个重要的里程碑是文献库形成过程中的结构显性和人工智能发展中的知识标识，这两种过程形成的不同目的和模式的信息显性结构，都是专门进行的、对着一个相对复杂的信息集合和任务集合而开展的信息结构显性。

第六个阶段是以显性完备信息结构为基础的自主发展的信息空间的形

成，这个阶段在起步阶段，还需要几个十年的时间才会真正独立存在于地球文明中。在这个阶段，全域显性完备信息结构与各个信息域、主题域的完备信息结构基本成型，以此为基础，形成了在感知、传输、处理、存储、内容、执行、系统等信息技术支撑下的技术架构，以这一技术架构为基础，连接各类信息集合，形成不依赖人的从感知获取到学习、建构的显性完备信息结构的发展体系，以这个体系为核心，形成独立的信息空间。图 4.2 简要解释了信息增长六个阶段的作用及相互之间的关系。

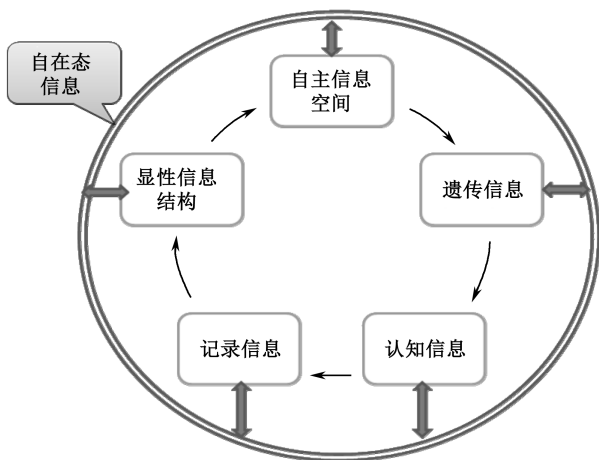


图 4.2 信息增长的阶段和相互关系

图中每一个阶段都有独立的含义和增长的方向和过程，每个阶段又存在依赖关系，不同阶段之间又存在相互的影响。自在态信息是信息增长的起点，是所有阶段发展的基础，也是信息增长质量的最终参照系。遗传信息在自在态信息的刺激下发生，遗传信息不断进化促使生物体认知功能不断完善，直到形成人类高度发达的认知神经系统。认知信息依赖遗传基因生成的认知功能，认知功能的进化产生了信息可以外化的共有外壳，使信息的增长摆脱了时间、空间、主体的约束，推动了记录信息的增长。记录信息持续增长，必然形成庞大的记录信息集合，造成利用的困难。记录信息结构显性随着记录信息增长而发展，对信息结构直接而有力的推动是智能化进程，信息结构显

性是智能系统的必要功能。独立自主的信息空间依赖于信息集合的增长和信息结构显性的进程，在更加先进和体系化的信息技术支撑下，在先行信息结构逐步完备的前提下，自主的信息空间必将形成，并为其他信息增长要素提供全面的支撑，走上信息增长的新阶段。

走向独立信息空间的增长历程中，这六个环节缺一不可，而且只有前一个环节取得足够的进展，后一个环节才能发生和发展，但后面的环节也会对前面的环节发展产生重大影响，本章后续部分将对此作更为深入的分析。

4.2 自有态信息的增长

迄今为止，介入信息发展的自有态信息有两种存在形态，即遗传信息和认知信息，可以预见的还有一种存在形态，即存在计算机信息系统中，且使用独立的、别的系统或生物体不能辨识的外壳。自有态信息的增长也包括量的增长和质的增长两个部分，本节将按三种形态、两个部分展开。

4.2.1 遗传信息的增长

遗传信息（在本节中，没有特殊注明的，遗传信息都是指自有态遗传信息，而不是其对应的认知或记录信息）是除开自在态信息之外历史最悠久的历史形态，从生命的一开始产生一直到所有生命毁灭之前，都是它的存续和发展期。即使从地球文明而言，今天我们还不能确切地预见这个事件发生的时间表，但应该晚于认知信息。因此，从存续时间看，有可能认知信息是自有态信息中相对较短的。

遗传信息的增长就是一部生命发展史，生物学和分子生物学对此有充分的研究²。生物物种的生生灭灭，直接导致了遗传信息的增长和减少，遗传信

2 参见基础生命科学（第二版），吴庆余编著，高等教育出版社，2006 年和分子生物学（原书第五版），[美]Robert F. Weaver 著，郑用琰、马纪、李玉花、罗杰等译，科学出版社，2013。

息究竟是多少,如何度量,到目前为止的研究还不能回答这个问题。人类基因组计划取得了巨大的成功,但只是编码了2%的基因,还有98%的基因没有编码,也就是还没有全面掌握其机理,但这是分子生物学家的的工作,不是本书的研究内容,本书关注的是遗传基因进化的哪些环节对信息增长有重要作用,遗传基因的研究又如何加快这些环节的基因的进化,对基因这类自在态信息的增长又产生了什么作用。

遗传信息的产生是信息发展史中关键的第一步,因为在它的长期演进中,形成了生物体的认知功能,特别是最终集成在人类身上的认知功能,推动信息增长进入一个新的阶段。约35亿年发展,遗传基因从生命没有产生前的0发展到人类拥有的6万个。这是一个不间断的过程,人类6万个遗传基因中包含了許多其他生物及早期形成的基因。从理论上来说,不能说哪个基因决定的生物体功能对认知神经系统的发展不重要,对认知信息的形成和发展不重要,但有几个关键环节对认知功能的发展特别重要。首先是神经系统功能基因的产生和发展。大约从5亿多年前开始软体动物基因中就包含了简单的神经系统,大部分无脊椎动物都含有神经系统的基因;到3亿多年前开始,脊椎动物开始繁衍,神经系统功能的基因得到进一步强化,而哺乳动物则发展了最为丰富的神经系统遗传基因。从生物学的角度看,自然选择原则促使生物体从比较简单的生存环境到各类天敌的增加,自然环境变化的剧烈,导致感知信息并能及时准确处置的功能性基因发展得到鼓励;从信息发展的角度看,则是自在态信息持续作用于生物体,促使遗传基因发生变异,而神经系统高级化的突变得得到自在态信息的持续鼓励,成为相对低级神经系统基因的替代基因,也就是自在态信息在自在态信息的刺激下进化。

其次是知觉功能的精细化。没有神经系统的生物就有知觉,知觉与生命相伴而生。发展到哺乳动物,知觉系统已经基本完备,但精细度或信息感知、处理的能力还不足。这种不足产生的后果,不是应对环境变化的能力不足,就是在竞争中告败,由信息承载的外部环境压力推动知觉功能持续朝精细化方向发展。例如视觉功能的进化,人类能看到彩色世界就经历了一个漫长而复杂的过程。在哺乳动物中,只有灵长类动物具有三色觉,其余非灵长类哺乳动物只有二色觉,个别夜间活动的只有一个色觉,当然个别鸟类具有四色

觉。灵长类动物并不是一开始就有三色觉，而是某个灵长类动物的色素基因在复制时发生突变，产生了不同的色素基因，而且这个突变在长期的自然选择中不仅保留下来，还成为灵长类动物色素基因的主导基因，人类作为灵长类的后裔，才拥有了三色觉³。

第三是人类的智力发展也依赖于遗传基因的变异形成新的功能，这里不对这一结论做详细的分子生物学介绍，只从两个直观的例子作简单的说明。一是人类发展史研究中的一个重大发现，就是找到了生活在公元前几十万年到 3 万年的尼安德特人的崛起和消亡之谜。在这长达数十万年的时期中，尼安德特人身体强壮，打猎本领、工具制作水平高于同期其他直立人，一度具有很强的竞争力。但是在与晚期智人的竞争中，很快消亡得无影无踪。为什么？研究发现尼安德特人与晚期智人的人体结构中有一个重大不同，那就是尼安德特人的喉部，包括声带，在嗓子里的位置都比较高，这一特征限制了利用咽部肌肉来调节声带发声的功能，导致尼安德特人发音能力比晚期智人差。而晚期智人的肺部有充分的弯曲，语言表达也由此丰富化、实用化，从而提升了大脑的能力，适应当时社会生活复杂化带来的交流功能的需求，晚期智人因此取代了尼安德特人，在不太长的时间内，成为唯一生存下来的人种⁴。二是生物学的研究说明，从直立人到现代人，认知系统的主要部分功能均发生了重大变化，而这些变化都是通过基因的进化而实现的。图 4.3 形象地说明了这一点，这一不同，不仅是形成大脑和大脑功能的基因在进化，而且相关的从知觉功能到整个中枢神经系统的功能都在相应进化，才达到今天人类智能水平。

从信息增长看，遗传信息的主要作用就是提升认知功能的遗传基因的增长。从发展看，今天人类大脑依然不能适应外部环境对认知功能的需要，更不能适应信息（含体系化知识）的爆炸式增长带来的获取、学习、判断和行

3 参见“进化里程碑：人类看到了彩色的世界” Gerald H.Jacobs and Jeremy Nathans, 姜世香、杨雄里译，载生命与进化，《环球科学》杂志社编，电子工业出版社，2011 年，第 86-93 页。

4 全球通史 1 人类的黎明，美国时代生活编辑部著，赵沛林译，吉林文史出版社，2011 年，第 36-70 页。

为的需要。人类作为地球上最高级的智能生物体，基于遗传基因的认知功能必然要继续发展。如何发展？有两条路径，一条是继续沿着基因突变、自然选择的方向，另一条是运用基因工程技术，加快变异的优化和稳定。第一条路是稳健的，几十亿年的进化就是这样过来的，但面对汹涌的信息爆炸，人类可能在认知能力的竞争中输掉，因为功能的进化之路太漫长了。第二条路是否可行，还有至少两大难题，一是技术上的，如上所述，人类基因组工程只编码了 2% 的基因，而没有编码，曾被称为垃圾基因的部分，已经证明至少 80% 在人类生命过程中是活动的，能不能理想地按照设定的路径优化认知基因的进化，在技术上还是不能证明的。二是伦理上的，社会是否接受这样的改变，能接受什么程度的改变，同样是没有定论的。我们希望在这两个方面都取得突破，使遗传信息的增长有一个历史性的突破，从而使人类在未来地球文明的发展竞争中保持优势的时间长一些。

从信息增长的角度看遗传信息增长的度量，可与遗传信息测序的度量一致，基本计量单位是碱基数 (b)⁵。

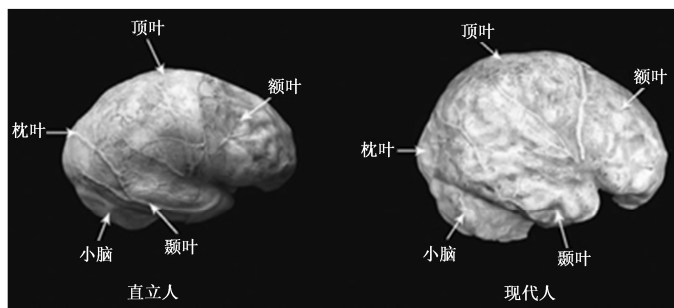


图 4.3 从直立人到现代人大脑形态的变化⁶

5 分子生物学（原书第五版），[美]Robert F. Weaver 著，郑用琰、马纪、李玉花、罗杰等译，科学出版社，2013 年，第 24 和 25 章。

6 图片来源：<http://szb.qxkj.com/html/2011-07/15/>。

4.2.2 认知信息的增长

认知信息的产生和发展是信息增长的第三个台阶，它既依赖遗传信息的增长，也受到它的产物——记录信息的影响，甚至还会受到第五个台阶——显性结构发展的影响。认知信息增长有两类主体，即生物体和非生物体，非生物体认知信息增长在 4.2.3 讨论，这里只讨论生物体，主要是人的认知信息增长。认知信息增长模型见图 4.4(避免重复，本图含非生物体认知信息增长)。

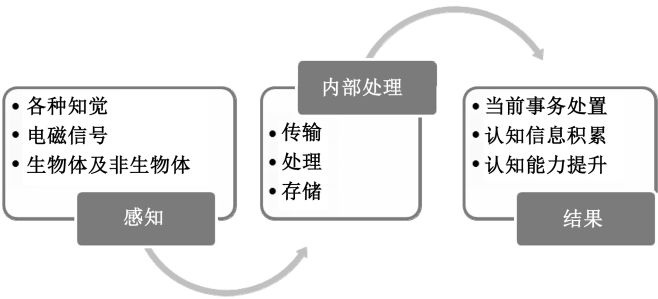


图 4.4 认知信息增长三阶段模型

如图 4.4 所示，信息增长的第一个阶段是感知，即获取信息。对于生物体，感知主要通过知觉器官，视觉、听觉、触觉、味觉、嗅觉；对于非生物体信息系统主要通过各种传感器，接受不同的电磁信号；需要指出的是，生物体通过电磁信号获取信息已经在一定范围内成功，大脑认知信息与电磁信号的转换也已经初步实现。第二个阶段是生物体内或非生物信息系统内的各类信息处理功能。一是传输功能，生物体将感知的信息通过中枢神经系统传输到大脑，经由大脑处理后再传输到执行的器官，书写、讲话、跺脚、挥手、皱眉，等等；信息系统则经由系统的传输信道执行。二是内部的各种信息处理，如信息系统的转换功能，将接收到的信号转换成系统可辨识的信息；如人将感知到的信息与当前事务连接起来，进行分析判断，将接收到的信息进行归类、分解、组合，融合到已有的大脑存储结构中，将经过分析、判断的信息存储起来，方便以后的使用等。三是存储功能，存储之前是需要处理的，但存储本身也是一个重要的功能。如存储的布局，既要满足存储的需要，还

要满足使用的需要；既要考虑长期存储，又要考虑传输和存储量的能力如何保证在一个时期后，存储的信息不变。在存储功能上，生物体主要考虑的是存储的结构及长期记忆与处理功能的关系，而非生物体主要的难度在于信息的结构，如何构成与应用一致的信息结构是关键。认知信息增长的第三个阶段是结果，认知过程产生了什么。有三个主要结果，一是认知信息的积累在认知过程中持续增长，过程不断，增长不断。二是认知功能的生长，生物体认知功能增长体现在对遗传信息的引导性反馈和已有功能的熟练和性能提升；非生物体信息系统认知功能增长在概念上与生物体一样，具体实现的技术方法和路径则完全不同。三是当前事务处理，从生物体感知信息或信息系统获取信息的角度看，这是第一位的结果，但对于信息增长，排在第三位。事务这个词在这里是其最广泛含义上的使用，除了工作、学习、生活、娱乐全部包括在内，还包括了情绪、聊天、性格等方面的内容。

度量认知信息增长，有两个主要的指标，一是个体认知信息的水平和数量，二是在给定信息域内同类生物体的认知信息的整体水平和总量。由于人类在认知能力和记录信息能力上远远超过其他生物物种，所以在以后的讨论中只讨论人类认知信息的增长。

如图 4.5 所示，决定一个个体认知信息的水平和数量，有五个要素：一是遗传基因和生理能力决定的认知能力，二是所处的生存环境，三是交流对象，四是记录信息的水平和可获得性，五是从事的职业。

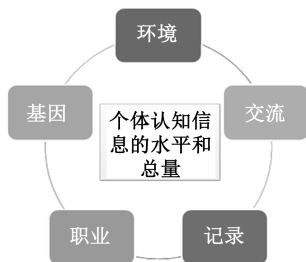


图 4.5 认知信息增长五要素

遗传基因和生理能力决定的认知能力是个体认知水平和总量的关键要

素，一个先天色盲不可能精细辨识色彩，一个后天失聪者没有听觉，一个先天智障不能学习高深的数理知识，等等。但这只是少数人，大部分人具备基本相同的遗传和生理能力，自从语言产生到现在的几万年时间里，现代人的认知基因没有发生突变，或即使发生了突变，也没有成为改变人类认知基因的自然选择。大部分人，认知水平和认知信息总量依赖于后面四个环节。

生存环境对个体的认知水平发展和认知信息积累具有举足轻重的作用。从整体看，是与经济社会发展水平一致，史前社会的人，不可能认识文字，不可能知道铁制农具；工业革命以前，就是在英伦三岛和欧洲大陆，也不可能认识蒸汽机、铁路、汽车；信息革命以前的人，不可能认识计算机和互联网的作用和威力。这是一概而论，工业革命以后很长时间，很多地区的人没有机会认识汽车、铁路和飞机，到现在，还有一些人游离在工业文明之外。认知水平和积累的认知信息，一定是与生存的环境紧密相关。一个孩子从出生到说出最简单的字，大约需要一年时间，遗传信息给予了功能，但需要父母和周边人长时间的配合教育，才能把肌肉、声带与语音结合起来；一个孩子在狼群中长大，尽管有语言功能，但不会说话，而回到人群中，又会慢慢学会说话。生存环境对认知能力的发展具有重要作用。

交流在人类认知能力和认知信息积累过程中扮演了如同化学反应中的催化剂角色。晚期智障人在众多人种竞争中胜出，其语言功能优势是主要的原因，语言是个体认知信息积累在交流中加速的工具，个体积累和群体积累结合起来，尽管仅用语言，涉及的群体还很小，但也有巨大的作用，“三个臭皮匠，胜过诸葛亮”在一定意义上解释了这个意思。交流对认知信息水平提升和积累，并不只依赖具有交流环境，和谁交流同样十分重要。交流对象的互补性、增长性都是重要的因素，“听君一席话，胜读十年书”就是交流对象重要性的脚注。交流的发展还依赖于信息记录和传输技术的发展，记录技术的发展拓展了交流的时空范畴，传输技术的发展扩展了交流圈，互联网更将交流的平台拓展到了全球，交流的内容也达到了前所未有的丰富，而且是持续加速发展，互联网已经成为个体认知信息增长最重要的平台。

文字和文献诞生之后，记录信息逐步发展为交流的主渠道，也是个体认知信息增长的主渠道。一个历史时期的记录信息实际上代表了 this 历史阶段

人类的认知水平，即科学技术的发展水平，一个时代的人类有一个时代的认知起点。个体能否有条件获取这样的信息，是不是有能力吸取这样的信息，是个体认知能力和认知信息积累的重要条件。平均受教育时间不断提升，既是社会发展水平提升的标志，也是记录信息系统性增加的结果，更是个体认知能力的信息（知识）基础。文盲、初等教育、中等教育、大学本科、博士研究生对记录信息的获取基础在完全不同的起跑线上。

基础能力是个体认知信息积累提升的一个条件，记录信息可获得性是另一个条件。可获得性是技术问题，也是经济问题，在一定范畴内也是政治文化问题。技术问题有两个侧面，一是社会记录信息和传输信息的能力，将在本节下一部分讨论；二是外壳的一致性及转换能力的变化，也就是处于什么信息域及不同信息域的转换，外壳不一致，不能转换，获得了记录信息也不能利用。经济问题也有两个侧面，首先，从社会的角度是社会发展水平决定的记录信息总量及公共服务的能力，国家或地区对记录信息收藏并向公众开放的能力，决定了整体的可获得性；其次是个体的支付能力，在不同的具体环境下，支付能力的构成并不相同，包括购买、旅行、复制等费用的支付能力。

从事的职业也是影响个体认知信息积累的重要因素，但不影响社会整体认知信息的积累。职业在社会中的位置，一个人的职业决定了什么社会信息流向这一主体。职业的细分进一步决定了更加具体的信息流向，如一个生物学的教授更多地积累了生物学相关的信息，而不是同一个大学教数学的教授在数学相关信息的积累。即使是大学同一个专业毕业，从事研究的岗位和从事工程、商务等岗位，各自积累的信息也大相径庭。

从信息增长的全局和全过程看，更需要分析群体的认知信息积累的能力和水平。自从语言文字产生之后，个体认知能力和认知信息与群体能力相互影响，加快了认知信息积累的速度，提升了认知信息积累的质量，一个信息域、一个国家或地区的认知能力和认知信息积累依赖于个体，体现在群体，留存于载体。这三者的互动构成了一个个历史时代的认知信息增长的实践，成为推动信息增长进入新台阶的动力。

4.2.3 非生物自有态信息的增长

在第2章已经提出,除了遗传信息和认知信息外还有第三类自有态信息。作为自有态信息,必须是一个主体拥有并能理解和利用的信息集合,而且只有这个主体可以理解和利用,只有这个主体可以使这个信息集合增长或发挥作用。计算机信息系统中存在该系统拥有并可以理解和利用的信息集合,但不是私有的,更不是专有的。当计算机信息系统经过发展,可以自主获取信息,自主学习和发展,并自我形成与人类的符号和概念体系不同的信息外壳,便成为第三类自有态信息的主体,本质上就是非生物体的认知信息。

目前还没有形成真正意义上的第三类自有态信息及其主体,但一些迹象或雏形正在显现。一是第3章讨论过的“沃森”系统,不管是用于“Jeopardy”智力游戏的系统还是注册为学习数学、英语的本科生系统,只要给予系统更多的获取信息、自学习和自组织功能,对获取的信息系统使用适合处理和理解的新外壳体系,与人类的交互通过转换,都有可能朝着第三类主体发展。二是在谷歌、百度等企业的一些智能系统,已经具备比较大的自我获取信息和学习的功能,工程师若放弃主控权,并允许采用不同的外壳,也可能朝着第三类主体发展。

第三类主体拥有的第三类自有态信息是信息增长第六阶段的重要组成部分,是信息增长的必然要求,在不太长的时间内,我们会看到它的产生。认知能力和认知信息的积累,将是智能竞争的关键环节。谁在与人类在认知能力上竞争呢?显然不是其他生物,其他生物这方面的进化已经被人类远远甩在后头,自然进化过程不可能让它超越人类。而人类的认知能力和认知信息积累能力正受到生理功能约束,面对日益增长的信息和处理信息能力增长的压力,但生物体的功能进化需要以千年、甚至万年为单位。由于信息技术演进的规律,计算机信息系统没有这个约束,所以,能超越人类的将是拥有第三类自有态信息的主体。

认知信息增长的度量有两种方式,一是对一个认知主体长期记忆的信息,度量该主体认知信息的总体数量和质量;二是对一个认知主体一个特定时刻认知信息总量,度量该主体认知能力和需求。度量的理论方法是映射到对应

的显性信息结构，基本计量单位是 mit。目前基本不具备度量的条件，但随着第三类自有态信息的产生和发展，度量的可能性和必要性都会大幅度上升。

4.3 记录信息的增长

在信息增长的链环中，记录信息处于中间，起着承前启后的作用。对于前面三个阶段，记录信息是遗传信息和认知信息增长的结果，是自在信息历经数十亿年时间，产生了在信息结构意义上一致的、不依赖具体生物体的新的客观存在，是信息进化过程中质的飞跃。记录信息的增长要回答记录什么、记录的载体和外壳是什么、记录增长的必要条件，记录增长的度量等问题，本节分别讨论这些问题。

4.3.1 信息记录的过程

记录信息的形成有两条主要路径，一条是经由人或非生物信息系统的认知过程，将认知信息转换成记录信息，另一条是直接将自在态信息转换为记录信息。信息记录及发展的条件和过程如图 4.6 所示，下面分别予以讨论。



图 4.6 信息记录的条件和过程示意图

无论是上述哪一条记录路径，信息记录必须有信息源。信息源有三类，一是自在态信息，是传感器、摄录设备记录信息的对象；二是认知信息，把认知信息中的一部分转换成记录信息，文学艺术、科学文献、工程文档都是这一类信息源的转换品；三是记录信息，记录信息经由一个新的记录信息产生过程变成内容不变、载体与/或外壳变化的新记录信息形态：如电视新闻的

文字稿，会议上照本宣科地报告录音，或者是通过扫描设备将模拟的记录信息转换成数字型记录信息。

谁在记录信息，这就是主体。主体就是两类，一是人类，二是信息系统，到目前为止，还没有独立于人类的信息系统，因此，第二类也可以看作是人类控制下的信息系统。但在实际记录的场景中，存在大量不同的原因，在这里，不仅要回答谁在记录，还要回答记录什么。

首先是记录认知信息。把自己看到的、想到的、总结出来的信息中要告诉别人的，或有必要记录下来防止遗忘、用作证据的，等等，都成为记录信息的原因。在文字产生以后几千年的时间里，奴隶、平民、农民几乎没有能力、时间或财富来记录属于自己的认知信息，即使到了工业革命之后最近的200多年，市民也是很难记录并保存自己的认知信息，保留下来的文献中这样的内容少之又少。到了互联网普及，记录信息数字化之后，记录信息不再是奢侈品，数字相机、智能手机成为记录日常生活信息的主要工具，社交网站及各种网络又提供了方便廉价的海量信息存储，记录普通人日常生活和认知的信息这个记录信息发展的动因才得以实现。

其次是记录人类社会发生的事件。记录信息在纸张和印刷术发明之前，是昂贵的奢侈品，早期在龟甲、泥板、青铜器等载体上记录的信息，主要是部落或后期王宫贵族的专利，记录的也主要是部落、王室关于祭祀、战争、继承等的大事。在所有的记录信息中，统治者自身的或关注的社会事件记录是主要的部分。在社会事件中，记录比较系统完整的另一种类型是科技知识。记录信息一直与科技知识的发展同步，反映了各个时期科技的进步。到了信息技术高度发展的今天，社会事件的记录范围有了很大的拓展，从针对特定空间的记录，如道路、小区、桥梁、车间等的监控，到针对特定活动的记录，如会议、演出、展览等的全过程记录，都在扩展和深化。

第三是记录自然界发生的事件。在历史文献中，地震、洪水等重大自然灾害的记录不少。随着经济社会和技术的发展，记录自然界发生的事件的需求不断提升，记录的内容不断丰富。从动物到植物，从地下、地面、地表到空间、太空，从气象变化到太阳系、银河系甚至宇宙的变化，从温度、湿度、气压到空气、水、土壤中的成分变化，记录的范围不断扩展，质量持续提升。

记录要有外壳作为信息的载体。在记录信息这个特定的范畴内，外壳主要是符号体系和概念体系。符号体系是语言、文字以及可以借由语言、文字解释的图像、图片。概念体系是由词和词间关系构成的语义逻辑，规范、注解一种语言文字的使用。人类社会是多语种的，因此符号体系是多元的。概念体系在跨语种的交流过程中不断同化，总体是大同小不同，为记录信息跨信息域的转换提供了条件。以0和1表示的信息，0和1是符号体系，不是概念体系。符号体系对应什么语言和文字，是规定的转换标准确定的，它没有解决不同信息域转换的问题，也没有与0和1挂钩的概念体系。

在这里需要对语言这种符号体系作一个特别的说明。在音频记录信息中，语言是记录的符号；在人与人交流的环境中，语言是交流各方能理解的符号，是共有，不是私有；语言和文字有紧密的直接关联，一对文字和语言表达相同的概念；但语言又是认知信息的直接构成部分，连发音都是认知能力的构成部分。如果没用电磁信号记录下来，语言是不是记录信息的一个符号体系？一场报告会，不能说有录音的语言就是记录信息，而完全相同的散布与会场的其他部分就不是记录信息的符号体系，所以从记录信息的符号体系看，语言是一个重要的符号体系。

载体是记录信息存在的条件，而不同的载体又由不同的技术来支持。从第一段以文字形式记录的信息产生以来，经历了自然载体（泥板、龟甲、竹简、青铜器、丝帛等）、纸张、电磁信号等载体的发展，经历了刻画、书写、泥板印刷、活字印刷、机器印刷、复印、扫描、数字出版、网络出版等记录模式，经历了人送、马车、公路、铁路、水运、空运、网络等信息传输模式，经历了模拟为主到数字为主的转变。

历史证明，在社会的各个发展领域，记录信息数量的增长总是放在很高的优先级上。提高载体记录能力的技术也放在很高的优先级上。

记录信息的增长，无论是数量还是质量，数字形式的出现是一个质变。数字形式首先出现在计算机领域，计算机诞生之前，就完成了进入计算机系统信息的数字化技术准备，布尔逻辑和芯片、软件、输出设备全部能实现二进制计算。通信系统、摄录设备等领域随后实现了从模拟走向数字的转变。模拟到数字转换的另一个重要领域是各类业务系统，这里的业务系统包括制

造、服务、管理、农业等各个经济和社会发展领域的信息系统。各种自动化系统逐步从模拟的感知转为数字化处理，各类仪器仪表快速从模拟信号收集和处理的数字信号。信息记录和处理的数字化与信息传输的数字化融合到一起，构成了互联网条件下信息增长和利用的新环境，这个环境使得记录信息进入爆发式增长期，信息利用得到了融合的成效。一些研究认为，最近两年生成的记录信息是人类历史除开这两年的总和，其实构成这记录信息数量的主要是自在态信息直接转换成记录信息，而不是认知过程的产物。

各种记录信息在产生后，为了保存和利用，必然会形成一个信息实体的集合。在几千年的历史发展过程中，记录信息集合的规模、属性、形态发生了重大的变化，总的趋势是类型越来越多、规模越来越大、获取越来越方便、成本越来越低。

记录信息集合大体可以归为四个系列。第一个系列是国家的信息集合，或者称之为政府信息，由两个主要部分组成，一类是国家事务的记录，称之为档案，另一类是国家收集的记录信息，称之为国家藏书，国家图书馆及公共图书馆就属此类。记录信息刚诞生的时候，部落或王室小心翼翼地把记录有信息的泥板、龟甲归集在一个安全的地方，可能这个名字都没有。以后书写的文献产生，记录内容增多，才开始分门别类，形成不同的国家收藏记录信息集合。因此，这一系列的信息集合类型变化不大，根据收集范畴的不同，档案有所区分，如国家和地方、国防与经济等；国家藏书也根据内容的不同有所区分，如主要收集科技文献的称为情报所，专门收集医学文献的称为国家医学图书馆，等等。类型虽然变化不大，但数量和规模却在不断增长，增长的速度与规模与经济发展和人口的增长成正比。由于数量和规模的扩展，利用也更加方便，到互联网时代，数字图书馆实现了足不出户的公民服务。规模增长最快的，在记录信息中占有重要位置的是档案这一类政府信息的变化。现代政府治理架构形成之后，政府的管理、服务、决策对信息的依赖程度越来越高，政府收集和生成、利用的信息越来越多，各级、各个政府组成部门及其关联机构的记录信息集合成为政府记录信息中关注度最高的信息集合，是开放数据、政府信息公开的主要对象。

第二个系列是机构发展过程中形成的信息集合。这里机构是指各种法人

或非法人组织。机构信息集合早已有之，在工业企业出现之前，一些商贸机构积累了自身业务相关的记录信息，寺庙、修道院这类机构产生和保留着记录信息，稍后产生的教育、医疗、科研机构，记录信息就是这些机构生存和发展的命脉，工业革命之后，企业成为经济的基本单元，企业记录信息在社会记录信息中的位置和比重日益增加。随着业务延续和发展，传感、数字和网络技术的进步，机构拥有的记录信息集合规模将持续增长，成为社会中最重要记录信息类型。

第三个系列是属于个人或个人构建的记录信息集合。个人的信息集合已经存在超过千年的历史，中国古代的藏书楼多为个人所有，但那个时候，只有少数富贵而又钟情藏书的人才有这种可能，数量少之又少。到印刷品成为社会记录信息的主要生产方式，社会有钱人数量的增加，个人藏书逐渐成为文化人的追求，个人记录信息集合的数量快速增加，但每个集合的数量有限。数字技术和网络的普及，获取信息和存储信息十分方便廉价，几十个 G 的网盘、几十个 T 的移动硬盘，都是一般中产阶级或工薪阶层所能支付的。因此，个人记录信息集合不仅在拥有者的数量，而且每个信息集合的规模也在持续增长。从记录信息集合数量看，个人拥有的已经超过政府和机构，这个差距还会快速加大。

第四个系列是互联网上的记录信息。互联网是一个平台，网上的信息集合不仅包括前三个系列，还有新的属性。政府、机构和个人的信息集合不同程度存在于互联网上，有的机构和个人的记录信息集合甚至全部都在互联网上。互联网上的信息又产生了一种新的属性，那就是放弃知识产权或可以免费获取并复制的记录信息成为一种全社会共有的信息，可能源自上面三个系列，但其属性不再是这三个类型的信息集合所能包含的。这种属性不仅为新的信息集合的产生创造了条件，更为不受限制的信息利用提供了基础。

上面讨论了迄今为止记录信息产生和发展的过程，没有讨论未来可能产生的新模式，那就是非生物信息系统记录信息的自我生长过程。这样的系统根据自身的意愿，通过连接在系统上的输入（感知，或直接的可识别其他数字信息）经由自身具备的信息加工、组织、管理功能，形成新的记录信息集合，并不断增长。这一过程在某些系统已经初步实现，只是或多或少存在人

的干预环节。在发展的过程中，这样的环节越变越少，经过一段时间后，新的记录信息集合产生模式就形成了，那就是记录信息形成的自动化生产线。

4.3.2 记录信息增长模型

从 4.3.1 对记录信息增长过程的讨论，可以分析出影响记录信息增长的四个关键环节，即主体、技术、社会环境、外壳。构成主体的认知信息表达的意愿和对信息的需求为出发点，通过外壳、技术和社会环境构成基本的记录能力，最终形成众多的信息集合。这就构成了如图 4.7 所示的记录信息增长一般模型。

如图 4.7 所示，决定记录信息集合增长主要有八个因素，分别是主体的数量、认知水平、记录能力、表达能力、传输能力、组织能力、处理能力、支付能力。下面分别讨论。

主体表达的欲望和获取信息的需求，是记录信息产生和发展的原动力，但并不是只要有了这样的动力就能取得一样的结果。在历史发展过程中，为什么有时候发展快、质量高，有时候则反之，或处于这两者之间的中间态，说明记录信息的增长是众多因素共同作用的结果。从主体的角度看，主要是具有记录能力的主体数量及这些主体的认知水平。

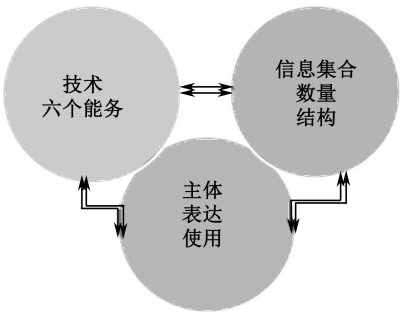


图 4.7 信息增长一般模型

显然，在同样的环境条件下，具有记录能力的主体数量越多，记录的信息越多。决定记录信息生成的数量，需要考虑多方面的因素，设具有信息记录能力的人的总数为 R_1 ，具有信息记录能力且从事记录的数量为 R_2 ，平均每个从事记录的人平均每年记录的信息量（载体量，以标准页计）为 nPB_1 ，当时的人口总数为 R ，设信息系统记录信息的点数为 S_1 ，平均每个端点年记录的信息量为 mPB_2 （载体量，以 byte 计），假定 $1PB_1=1000\text{byte}$ ， $1PB_2=1\text{byte}$ ，则我们有如下的定量分析指标：

有记录能力的人数与人口总数之比为 R_1/R 。有记录能力且从事记录的人数与人口总数之比为 R_2/R 。人记录的年信息总量为 $nPB_1 \times R_2$ 。系统端点记录的年信息总量为 $mPB_2 \times S_1$ 。年记录的信息总量为 $nPB_1 \times R_2 + mPB_2 \times S_1$ 。

比较上面的指标，纯粹从信息量看，若 $nPB_1 \times R_2 = mPB_2 \times S_1$ ，也就是人与系统记录的信息量相等，只要 $m \times S_1 = 1000n \times R_2$ ，即系统信息端点数与年平均记录信息量的积等于所有从事信息记录的人数与年平均记录信息量的积 1000 倍就行。从这个等式看，在人类历史数千年的时间里，人是记录信息形成的主体，但在网络技术普及，传感技术和信息记录技术数字化之后，两类主体在信息记录上展开了竞赛。随着传感器种类和功能不断增长，系统自动记录的信息增长十分快，尽管没有准确的统计，但在相当部分国家和地区，系统记录的信息总量超越了人类。

但这不是全部，而且从信息增长的目的看，还不是主要的方面。正如前面指出的，主体在记录信息增长中还有一个作用，那就是认知水平。信息增长不仅是为了解决面临的问题，更要加快形成完备信息显性结构，走向自我发展的信息空间，从这个维度看，主体的认知水平就是关键的因素。

主体的认知水平，从信息增长的目的看，就是能对显性信息结构的完备化做出贡献，就是能对还没有完备的部分通过认知过程一点点积累、一点点完善。完备的过程不是简单地重复或不重复地记录各类现象和认知，而是对已知新归纳、新结论，对未知的新发现，是显性信息结构中新关系、新属性、新功能、新信息点、新信息单元、新信息结构，或者是对已有的补充或纠错。当自主的具备认知功能的非生物智能体产生之前，这样的任务只能由人或人在人的指导下与信息系统共同完成。

信息增长的要素，除了主体的两个方面，还有技术和经济社会发展水平决定的六个能力。

记录能力是指将自在信息和认知信息转换为记录信息的能力。如图 4.8 信息记录的方式，主要有五种方式。

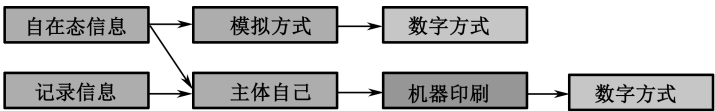


图 4.8 记录信息的主要方式

五种方式包含了三个台阶：模拟到数字、手工到机器、机器印刷到数字转换。每一个台阶都是信息记录的重大进步，而在每一个台阶中，还存在若干个小台阶，使用不同的技术方法。无论是模拟方式还是数字方式都直接将自在态信息转换为记录信息，感知的技术装备均具有重要作用。自在态信息直接记录能力构成中，技术装备和经济因素起着最关键的作用。认知信息记录能力构成中，机器印刷和数字方式都是生产率的提升，也就是量的增长，而认知信息转换成记录信息，还是依靠人自己。图中的主体自己有三个含义，首先是指人自己生产记录信息，即用刻写、书写的方式来生产记录信息；其次是指人类将认知信息到记录信息的转换过程；第三是指非生物信息系统将经过系统加工过的信息转换为信息记录信息。这一分析，指出了一个重要的主题，记录信息的生产实际上有两个性质不同的过程，一个是将如认知信息转换为记录信息，另一个则是将规定的自在态信息或已有的记录信息复制或大规模生产，这两种记录能力在信息增长中扮演不同的角色，但缺一不可。

表达能力是指自有信息的表述方式转换为共有信息表述方式的能力。有三种基本的转换类型，一是人类认知信息的转换，能力体现在语言、文字表述大脑认知信息的精确性、全面性、可理解性；二是系统感知的信息转换成系统可处理、可执行的信息，在数字技术成熟之后，就体现在模/数、数/模的转换能力；三是非生物信息系统自有的认知结果转换为与人共享的共有信息，也就是除人类语言文字之外的外壳形式转换成一种人类通用的语言文字的能力。表达能力的发展首先经历了一个缓慢的成熟过程，从语言的产生到成熟

的概念体系形成走了几千年，成熟的概念体系形成后，处于持续的新陈代谢过程，不断有新的语言文字产生，又有陈旧的淘汰掉。后两种模式是技术发展过程，相关技术进步是决定因素。

记录信息传输能力是指将记录信息从一个地方传输到另一个地方的能力。记录信息对传输能力的需求是一个双向的发展过程，记录信息的数量越多，需要该信息的人也越多，传输的要求也更高。在数字信息出现之前，信息的传输如同其他物质产品的传输，与社会交通运输能力同步。信息数字化之后，信息传输进入信息网络传输新阶段，因此，传输能力与记录能力有着多重相关性，既有正相关，又有负相关，记录能力推动传输能力的增长，传输能力的增长可能减少对记录的需求。

记录信息的组织能力是指记录信息集合通过何种方式组织能更有效地利用，整个社会采用何种模式，才能使记录信息的生产、传输、保管、利用的成本最低而效用最高。信息集合如何有效管理，不管是数字化的还是印刷文献集合，关键的问题是信息的结构化，按照主要用户的要求结构化是为使用服务的信息集合组织的主要方法。信息集合的组织方式，说到底就是一个结构化的问题，信息结构在多大颗粒度水平上显性的问题。

记录信息的处理能力是指满足记录信息形成、传输、组织、利用过程中的处理能力。在记录了信息形成的头几千年里，各个环节的处理能力都是人来完成，在计算机进入商用之后，信息技术在记录信息处理中扮演越来越重要的角色，激光照排、告别铅字，数据库管理系统、大规模数据中心信息资源管理系统等，都是记录信息处理能力的经典例子。

记录信息的支付能力是指从记录信息的形成到使用的全过程中，由谁来支付，支付能力与记录信息增长的关系。尽管支付能力一般地与经济社会发展水平成正比，由此得出记录信息形成到利用全过程发展与经济社会发展水平保持一致的结论，这个结论也与对不同发展水平的国家、地区记录信息的发展状态观察一致，与一个国家或地区历史发展过程的观察一致。但还有两点需要指出，一是在互联网普及之前，记录信息发展利用全流程支付的制度设计，特别是公共财政对面向全社会的记录信息收集利用支持力度，这对记录信息的发展具有举足轻重的影响；二是在互联网普及的环境下，公共财政

对互联网发展的支持力度和收费政策，采用与社会基础设施建设运行一致的政策，并最大限度地降低收费标准，鼓励机构和个人的使用，对数字化、网络化记录信息的增长有重大影响。

记录信息的负增长。前面主要讨论了记录信息正增长，但在记录信息发展的漫长历程中也发生了大量的负增长事件。所谓负增长就是由于人为的或自然的原因，一个信息域的部分或全部记录信息集合消亡了，一个信息域的重要主体或全部主体消亡了，甚至一个信息域信息增长的主客体都消亡了，这些事件的结果都是记录信息的负增长。可考证的记录信息负增长，不少于数百起，而不可考证的，尤其是记录信息产生早期的历史年代中，记录信息的负增长事件更多。不管有多少负增长的事件，整体上人类社会记录信息的增长并没有受这些事件而止步，从全局看，对增长速度的影响也是十分有限的。

记录信息是构成信息空间的主要组件，记录信息在信息增长的链环中居于重要位置，需要把握记录信息增长的规律、实现记录信息的有效、快速增长。

4.3.3 记录信息的度量

衡量记录信息增长的数量和质量，主要有三个指标，一是记录信息的数量，二是记录信息相对于记录对象的覆盖度，三是记录信息服务对象的满足度。一般而言，第三个指标是第一和第二个指标的函数，所以不做专门的讨论。第一个指标在 4.3.2 中已经分析，本节只讨论第二个指标。

理解这个指标，首先要讨论记录对象是什么。作为信息记录的对象，有三个大类，一是人类认知信息的记录，二是社会状态的记录，三是自然界状态的记录。

定义 4.2: 记录信息覆盖度。记录信息覆盖度是指已经记录的信息与实际需要记录的对象所包含的信息的比例。这里，记录对象实际存在是指认知、社会状态和自然状态三类，比例需要在三个维度确定，一是范围，二是颗粒度，三是时间跨度与连续性。下面分别针对这三类对象三个维度进行讨论。

首先讨论认知信息记录覆盖度。从记录覆盖度分析的角度看, 认知信息可以划分成三个类型, 分别是学术类、文学艺术类、日常事务(含工作、生活)类。其中学术类和文学艺术类都是创造性劳动的结晶, 作为形成这些认知信息的本人, 都是希望能够记录下来的, 如果发生已经形成而没有记录下来, 主要原因是记录成本过高或时间约束, 在一定的历史时期, 也曾为宗教、文化、政治等原因而不能记录, 在社会和技术双重进步的前提下, 这样的情景将逐渐消失。结论是, 在历史上确实发生过因经济、技术、成本、政治、文化、政治、法律等原因而使学术信息没有记录下来, 但正在逐渐消失。考虑到学术发展的延续性, 即使以前没有能够记录下来, 也在后来的研究、创作中得到补偿, 也就是说, 学术信息记录的覆盖度接近于完备。文学艺术创作的成功始终通过一种社会选择的功能, 有选择地记录, 尽管不能保证这种选择一定是准确的, 尤其是在一些特定的政治时期。但总体看, 这种选择机制还是有效的, 代表一个历史时期的文学艺术创作水准的作品能够记录下来。随着互联网和数字技术的发展, 信息记录和传播变得十分方便和便宜, 加上一些激励机制的负作用, 学术和文学艺术类信息的记录出现了鱼目混珠、泥沙俱下的局面, 在有些时候甚至将有价值的学术和艺术信息淹没在不怎么有价值, 甚至没有价值的信息中, 出现了超覆盖, 或者叫超记录。因此对学术和文学艺术类信息, 记录覆盖度(按定义 4.2, 已经记录的学术和文学艺术信息与需要记录的比)应保持在接近于 1, 或在 1 的左右小幅度摆动, 而不能远离 1。对于事务类认知信息的记录, 它的规律类似于社会和自然状态信息的记录, 下面一并讨论。

对于社会和自然状态信息记录的覆盖度, 两个变量中, 已经记录的信息只要设计专门的统计就可以确定, 但另一个变量, 需要恰当定义, 针对个人、机构、国家或地区、全球四个范畴分别讨论。

个人或家庭的有关情况的记录, 既可以作为个人认知信息的一部分, 也可以作为社会状态信息的一部分。如果这种记录只用于该个体或家庭, 那么覆盖度的度量由这个个体或家庭来承担, 这里不作讨论。但是个人的工作或生活信息、家庭的日常生活信息作为社会状态信息的一个必要组成部分时, 就应列入社会状态信息记录覆盖度中。

从机构的范畴看,根据机构性质或业务范畴的不同,有的机构只收集社会状态,有的只收集自然状态,有的两者兼收。因此,机构收集的社会或自然状态信息,是基于业务范畴和能力的。对于机构,上面一节论及的六个能力分成两类,即支付能力和其他五个能力。机构,不管是企业还是科研单位,对于业务开展需要记录了两类状态信息,只要支付能力能支持,一定充分使用另五种能力,使信息的记录满足需求。但是,只有极少数机构能够拥有足够的支付能力,其他机构只能在支付能力的约束下,对需要记录的两类状态信息进行筛选,确定记录的内容。在机构确定记录内容时,除了支付能力外,还有一个重要的能力是技术能力,如一个制造企业的自动化生产线,没有一个企业不希望通过更好更实用的传感器使自动化水平更高、质量控制更强,这样的時候,传感器技术就是瓶颈。在机构层面,信息记录的覆盖度高低,取决于机构领导的认识。一般可以判定为,如果机构领导认识到信息记录对机构业务开展的意义,信息记录的覆盖度在给定条件下,也是接近于1。

从国家或地区的范畴看(这里不包括个体和机构),需要记录的社会和自然状态信息至少由三个主要部分构成。一是社会突发事件,用于该事件的处置;二是日常社会管理和服務需要的相关社会状态和自然状态信息;三是各类研究需要的社会和自然状态信息。这三类信息又交叉重复,但各不相同。记录的过程一般与政府部门的职责分工一致,因此各个政府部门职责范畴内的两类状态信息记录的覆盖度也可以判定为接近于1。但是,无论是社会状态还是自然状态信息的记录,都存在超越部门职责的部分,都存在业务流程之外需要记录的信息。部门、机构自然叠加的总和并不总是最优的。例如气象、地震、水文、地形地貌、基础设施、建筑、经济社会发展的统计、城市交通等信息,都有固定的职责和流程负责记录,但此外还有很多需要记录的信息,如对地观察的频度和颗粒度、外太空观察、人群的集聚动态、人口流动动态,等等。国家或地区两类状态信息的记录,需要有一个整体的规划与分析,把记录什么,记录的范围、频度、颗粒度、指标体系标准等制度化、标准化,落实责任,才能满足当前和今后管理、服务、研究的需要。

在国家和地区之外全球范围信息的记录,主要是指自然状态信息记录的分工,因为许多重要的自然状态信息跨国界,公海、两极、洋流,以及国际组织需要开展的跨国研究项目涉及的在相关国家和地区还没有记录的两类状

态信息，需要专门的组织和规划。在特定的条件下，也包括社会状态的记录，如反洗钱、缉毒、反走私、反恐等。

4.4 信息结构的生长

定义 4.1 把信息结构的生长定义为质的信息增长，把隐性结构和显性结构、不同完备程度、不同信息域和主题域、新增信息结构已有结构的完备等内容均包含到信息增长的范畴中。本节开始，全面讨论信息结构的生长，重点在显性信息结构的生长。

4.4.1 信息结构生长的定义

定义 4.3: 信息结构生长。信息结构生长是指信息结构及其体系从形成到完备的发展过程。这个定义规定了三要素，一是明确信息生长是个过程，生长当动词用；二是既指一个信息结构的生长，也包含整个信息结构体系的生长；三是包括了信息结构生长的全过程，强调了生长的延续性。

在信息结构的定义中包含了隐性结构和显性结构。在隐性信息结构中，遗传信息的生长过程和如何继续生长以及隐性结构的显性化，是分子生物学研究的对象；认知信息结构的生长和显性化是认知科学，特别是认知神经科学研究的对象，在本章后边部分关于信息结构生长的讨论，集中在显性信息结构方面。

4.4.2 显性信息结构生长的一般模式

如图所示，显性信息结构生长的目的或方向是构建成覆盖所有信息域的显性完备信息结构。图 4.9 勾画了三条主线，中间一条是范畴的扩展，从一个信息点的属性及功能描述的初步形成作为起点，经由信息单元、信息结构、主题域、信息域一直到全域显性完备信息结构的形成。上面一条线是属性系列

的完善，有两个关键环节，一是从信息点描述的完善开始，经由信息单元、信息结构、主题域、信息域一直到全域的所有属性描述的完善；二是从信息点的索引，经由信息单元、信息结构、主题域、信息域一直到全域的所有属性系列和功能系列描述的索引建立并完善。下面一条线是功能系列的完善，同样需要从两个链环展开，一条链环是范畴，同样从信息点的功能形成开始，经由信息单元、信息结构、主题域、信息域一直到全域功能的持续完善；另一条链环是感知、连接、处理三个系列功能在各范畴层面及范畴之间的协同和完善。属性系列和功能系列的完善分别在 4.5 和 4.6 两节讨论。

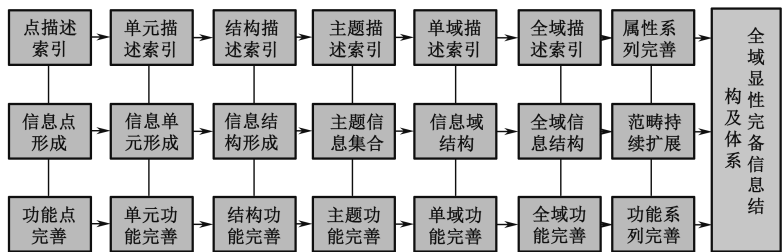


图 4.9 显性信息结构增长模式图

4.4.3 显性信息结构增长的基本规范

从图 4.9 看出，显性信息结构的发展是一个长期的过程，也是由不同主体、在不同地区并行发展的过程，是极其复杂规模庞大的任务，也是一个从小到大、从粗到细、从不完备到完备的自下而上的过程。这样的四个特点，决定了需要确立一系列原则、标准和规范来保证发展过程中的一致性、可用性。

但是，信息结构显性的特点、难点、需要规范的重点又是需要在发展中认识，相应的标准、规范、原则的形成也需要在发展中形成并得到检验。这样的特点，决定了信息结构显性，必须要有系统、严格、可操作的标准、规范和原则，同时，又必然要与发展过程同步发展。但如同显性信息结构一样，需要起步的规范，才能为以后发展过程中的完善打下基础。

既要在一开始就明确一些一般原则，以对显性信息结构发展过程进行规范，又要兼顾信息结构显性过程中一定会遇到的现在还不能预见的问题，原则需要有弹性。以下十条原则是一开始就要遵循的。

一是发展原则。任何可能导致显性信息结构不能走到全域完备、自我演进目标的原则都不应存在。

二是可控原则。任何可能由于显性信息结构的发展，对人类和地球文明的发展产生负面影响的原则或行动应得到抑制。

三是一致性原则。在一个信息结构中，或在不断拓展的更广泛的信息结构体系，直至全域显性信息结构体系，要保持必要的一致，以保证发展的延续和协调。必要的一致并不要求完全一致，而是要求能够朝着加快全域完备显性信息结构的方向发展，又能满足不同发展阶段和不同应用需求的显性信息结构的特殊要求。因此，这里的一致性主要是指标准、技术规范、平台等内容，而不是一个具体的信息单元或结构中的描述。

四是协同原则。要建立一系列的标准和技术规范，包括必要时建立信息结构显性工作的分享信息平台，使得不同时期、不同主体、不同主题域和信息域的信息结构显性成果能够相互参照，存在协同发展的机制。但协同机制不影响具体工作中的特殊要求。

五是动态调整原则。要有一个机制，能够客观评介各原则，并对各原则中不适合新情况、新要求或背离了该原则设定时的初衷，则应在审慎调整的原则下，经过严格的程序，及时进行调整。

六是有区别原则。当确定了原则、标准、技术规范后，一般而言，需要共同遵守，但信息结构显性的形成和完善在很多不同，甚至相悖的场景下进行，应允许不同场景下信息结构显性呈现的多样性，特别是为信息结构完备目标与为特定需求而开展的结构显性工作，允许不一致，允许出现相互矛盾的状态。

七是冗余原则。对应一个相同的客观存在，不同视角、不同场景进行信息结构的显性，一定产生大量的冗余。对各种原因产生的冗余，应该持包容吸纳的原则，即允许冗余的存在，能留则留，不是朝着最简全域显性信息结构的方向走，就不要冗余。

八是容错原则。这条原则和六、七两条原则相关，允许冗余、允许互相矛盾的描述并存，实际上就是容错，但容错原则更加直接。这条原则的确立，使显性结构按照预定的标准或技术规范进行验证或评价时，不因为部分人，甚至很多人给予不利的评价而被删除。当然，容错和保留也是要尺度的，需要在标准和技术规范中细化。

九是连接优先原则。这个原则是指显性信息结构各个基本构成单元间，不管这个构成单元是信息点中的一个具体描述，还是一个相当规模的信息结构，不管是属性描述还是功能实现，存在的联系要全面保证，甚至当前语义或功能上还没有连接，而根据某些证据可能今后会有连接的也应保留连接。连接时信息结构的核心要素。

十是功能优先原则。在属性系列和功能系列发展中，由于属性系列比功能系列行对容易，要防止功能跟不上属性描述发展步伐的状态。在认知信息功能发展过程中，先是通过基因实现了认知系统发展的必要功能，其次是在婴儿出生之后，首先是锻炼功能，然后才能进入语言、文字、概念、逻辑的学习，而形成生理功能进化了几十亿年，功能变成可用，也需要数年时间，然后才是经由学习过程的人类认知能力发挥。在显性信息结构发展中，如果属性描述超前，功能完善在后，则一定会出现不可进化的结构，而不是可以自我完善的结构。

标准是信息结构和信息结构体系完善的基本工具，应该在信息结构显性的发展实践中完善。但以下六个标准有必要在一开始就起基本的规范作用并具备动态完善的内在基础。

第一，结构标准。结构标准是指关于信息结构的体系和信息单元、信息结构的一般性要求。主要规范三个内容：一是关于一个信息域的显性信息结构体系架构的标准，该体系中信息结构的层次、分类及相互关系，整个体系的总体描述及要求等；二是关于一个信息结构的标准，规范其构成、格式，提出满足全体系架构和功能的要求等；三是关于一个信息单元的标准，规范构成、格式及与信息结构中其他单元间的关系。

第二，属性描述标准。对于一个信息单元，基本上有两类信息点，一类是关于所反映客观存在的属性的，一类是关于所反映客观存在的功能的，严

格地说,功能的信息点在描述时也呈现属性特征,但功能系列信息点在信息单元中的根本任务是能在符合性和重现性验证中,完成所承担的功能。属性描述标准要对每个信息单元中信息点描述的格式、属性描述中各种不同要素表述的次序、标识、基本要求做出规范,对信息结构和主题域、信息域的整体描述单元中的信息点描述做出类似的规定。

第三,信息结构体系各构件的接口标准。连接优先、能连则连是信息结构显性的主要原则,数量巨大,泛在的连接,甚至一个信息点存在成千上万的连接,应该在信息结构体系的各组成构件间建立接口标准,使连接规范、有效、简洁。应明确接口的类型及其连接方式和启动条件,明确不同语义关联的不同标识,明确不同功能连接的不同标识和协同规则,等等。

第四,不同层级的组合分解标准。在一个显性信息结构的体系中,在各个层级经常发生组合与分解,应该有一个标准,规范组合分解的过程及随后的新旧结构间的关系,规范的重点是新旧结构间各要素的遗传及改变,遗传如何规定标识和使用,改变后如何定义新要素与旧要素之间的关系。

第五,增删的标准。信息点、信息单元、信息结构均在发展过程中不断产生增删。组合与分解是一种类型的增删,而更多的是直接增删。增删的标准是为了规范增删行为,并为增删后的连接做出规定。根据冗余原则,对删除应有验证过程。

第六,验证标准。在冗余原则和连接优先原则下,信息结构的发展采取的是放宽约束,发展中自然选择的策略,其实,作为生命进化过程产物的遗传信息和认知信息也是采取这一策略。但在结构发展的几个步骤中还是需要把握成熟度。本书在讨论结构增长过程,很少用正确性作为规范行为的标准,而采用成熟度,因为在认知发展及显性信息结构发展过程中,首先要防止正确与否的误判,留待长期的使用中判定。当遇到下列两个过程,需要把握好成熟度。一是构建最简完备结构的所有构件,要经过严格的成熟度验证;二是为特定应用而构建的专用信息结构,要经过严格的成熟度验证。成熟度验证标准主要依据使用和增长过程中的正向激励,所谓正向激励就是被应用而没有负面评价。因此,在信息结构的每个构成组件中,应该有使用及负面评价的记录。

作为初始的技术规范至少应该包括功能体系的整体技术规范、感知功能技术规范、处理功能技术规范，要将前述原则和标准在技术实现中固定下来。

整体技术规范是指对一个信息域所有信息结构显性过程的功能体系及其实现给予全局规范。是指导各个信息结构对感知、连接、处理功能技术规范制定，并在各信息结构已有的技术规范基础上，明确从信息单元、信息结构到主题域、信息域等各个层次功能体系的基本技术要求，以实现处理能力描述的规范和实现的协同。

感知功能技术规范是指一个具体的信息结构中对感知功能描述和实现的技术性规定。这里感知功能是指一个信息结构中所有信息单元及其信息点从外部获取信息的总和，而信息单元、信息点的外部可能是该信息结构外，也可以是该信息结构内。视具体信息结构所反映的客观存在的不同，感知的技术、方法、内容、信息形态亦不同。从传感器感知到记录信息获取，从信息结构外部感知到内部交换，从任务接受到结果执行，都是感知功能规范要覆盖的。感知功能技术规范要提出同类感知功能的描述及重现的技术要求，不同类感知信息处理的一致性要求，信息点、信息单元、信息结构对感知功能的一致性以及与连接和处理的协同性要求。

连接功能技术规范是指一个具体的信息结构中对连接功能的描述和实现的技术性规定。这里连接功能是指一个信息结构中所有信息单元及其信息点内外连接的总和，既包括属性描述信息点间语义关系的连接，更包括感知、处理等功能实现和协同内含的连接。规范应对所有连接的描述格式和实现方式提出要求，应满足信息域整体技术规范对连接的一般规定。

处理功能技术规范是指一个具体的信息结构中对处理功能的描述和实现的技术性规定。这里处理功能是指一个信息结构中所有信息单元及其信息点需要的处理功能的总和。其实，信息结构中任何功能都需要处理，而所有的处理也需要连接和感知，所以信息结构的功能是密切相关的联合体，但需要有所区分，以明确特点，按类规范。因此，处理功能技术规范首先要对不同类的处理做出区分，然后分门别类提出描述和功能实现的要求，提出不同单元间相同功能的一致性，协同处理的要求及可实现性的规定。

4.5 显性信息结构完备的充要条件

信息增长的方向是形成全域显性完备结构，这个结构能为全域的各类主体利用。如何实现各个层次信息结构的完备，这是本章以下部分讨论的重点，首先是定义显性信息结构完备的充要条件，然后讨论充要条件如何实现。

定理 4.1: 显性信息结构完备的判定标准。称一个显性信息结构是完备的，则该信息结构在给定的范畴内把所有认知主体的认知成果和获取的记录态信息遵循信息结构的原则、标准和技术规范，以规范和简洁的形式表述出来。这一判定准则是不证自明的。显性信息结构完备与否就是认知主体能够达到的水平，将所有该客体的认知成果和记录态信息按规则构成显性信息结构，一定是完备的。

这里给定范畴是指信息结构所表述的客观存在；所有主体是指在该信息结构所在的信息域中，这些主体的认知成果总和就是对确定客体的全部认知成果，并不是指对这个客体有认知的全部主体；获取的记录信息是指与该完备信息结构有关，但并没有经由其他主体认知过程的原始记录态信息；原则、标准和技术规范是指这个信息域关于信息结构的相关规定，如果不存在，则参考前一节内容，自行制定；这里的简洁是指应该有一个显性结构是不包含冗余的，并不要求所有相关结构都是简洁的。

定理 4.2: 实现显性信息结构完备的充要条件。称一个完备显性信息结构是可实现的，则必存在至少一个认知主体，具备实现该信息结构完备的意愿、技术能力和支付能力，能够连接到该信息结构相关的所有认知主体，并可获得这些主体的所有认知成果和已感知的记录信息。这一定理也是不证自明的。

这里的认识主体不仅指一个人、一个机构或群体，也包括非生物主体，如具有这样能力的信息系统；技术能力包括认知能力和构建显性信息结构的能力；支付能力是指能承担整个过程的费用；连接既指直接连接，也指间接连接，间接连接中，还包括可证明即使连接也不能提供新的认知成果或记录

态信息的无连接；可获得的记录信息同定理 4.1 的说明。

定理 4.1 及定理 4.2 确定了完备显性信息结构的判断条件和实现条件，而其中，承担主体的形成和发展、一般假定的确定、与认知主体的通达性、成熟度验证及记录态信息增长迅猛与结构显性滞后的矛盾是关键环节，以下将逐一进行系统分析。

4.6 信息结构显性的承担主体

信息结构完备的首要条件是存在并愿意承担这一事务的主体。本节讨论主体的性质和动力、必须具备的能力和显性信息结构的属性。

4.6.1 主体的属性及动力

承担显性信息结构构建和完备的人、机构或系统就是主体。在主体是人还是计算机信息系统这个问题上，正在经历着变革。在记录态信息产生之后的数千年时间里，从事信息结构显性工作的，不管是档案、图书、科技文献还是专家系统和知识工程，都是由有知识的人，甚至是著名的学者来承担。但近十几年来，这个格局正在变化，计算机逐步分担人的部分工作，承担着如自动收集信息、自动分词、自动标引、自动文摘、自动索引、自动分类等事务，而这些事项与构建显性信息结构存在密切的关系。由于构建显性信息结构工作量巨大，有不少事项需要计算机辅助，成为类似于工业生产的机械化过程，机器承担部分工作，人承担大部分工作。随着计算智能的进展，信息结构显性工作规范性的增加，必将由计算机系统承担主要的工作量，而人承担的工作将逐渐减少，如同工业生产线从机械化到自动化。最终，信息结构显性的工作将基本由计算机系统承担，如同工业生产线从自动化到智能化一样，从智能技术发展趋势看，这一转变不需要很长时间。

比谁来承担更重要的是动力问题，谁愿意承担这项工作，或承担这项工作的动力从哪里来。同样，在过去的几千年里，类似性质的工作，主要有四

种动力。一是国家意志。统治者认为这是国家治理和发展的重要事项，有国家组织并提供经费支持，从有记录态信息到今天，这一动力始终扮演重要角色，在发展早期和一些特定领域，如关于国家安全，甚至是唯一角色，也积累了很多有益于一些领域显性信息结构形成的基础。二是个人或机构兴趣。出于私人或公益的目的，如中国的藏书楼，或欧洲的修道院，收集文献、组织管理，总体上对显性信息结构的形成有帮助，但在四种动力中贡献是最少的。三是研究的需要。科学研究，不管是科学还是技术，都需要从大量的相关文献中找到适合自己当前需求的那一部分，这通常会占去科技工作者超过半数的工作时间。完整、细颗粒度、结构良好的索引、分类、语义网络等是有效的工具，这就成为第三类动力，而且经常与政府支持科研的动力结合在一起，形成国家范围开放，部分全球开放的，精细组织的科技信息体系。四是企业利益、商业目的。大企业为了研发和管理的需要，收集了大量的相关信息，并为企业使用的效率而进行了适当的组织，这种组织的成果成为信息显性结构形成和完备的基础。更加重要的是基于互联网提供信息服务的企业，既收集了大量的信息，进行了适用的结构化整理，而且尽可能采用计算机辅助功能，为信息结构显性提供了很好的路径参考。

本质上，信息结构显性并走向完备是社会进步的必然要求，是全社会得益的事项，需要政府支持下的各方共同参与，而其中最重要的就是激励更多的主体参与。

4.6.2 主体的能力

承担显性信息结构完备工作，至少需要具备建构能力、信息获取能力、信息处理能力和支付能力。还有一些其他工作同样需要的一般性能力，如协调、管理、沟通等，这里就不讨论。

首先是建构能力。所谓建构能力是指关于构建、完善显性信息结构的基本能力，主要包括具备提出先验性一般假定，制定或遵循已有规定制定相应的原则、标准和技术规范，明确实施的基本方法和路径等。先验性一般假定在下一节专门讨论，这里只讨论后两者。对于制定显性信息结构建构的原则、

标准和技术规范在 4.4 已经做了比较系统的一般性讨论,但具体到一个确定范畴的信息结构,需要承担主体在一般原则下细化,这是信息结构显性化和完备的总纲,也是承担主体首要的技术能力。原则、标准和技术规范作用周期是信息结构显性从起始到完备的全过程,是决定质量和效率的基本规范。由于迄今为止,记录信息结构化都是针对特定目的的局部性工作,依据给定范畴的记录信息特征及发展趋势的估计,既顾及长远的方向和目标,又适合当前工作的进行,需要有远见地制定发展原则、工作标准和技术规范。对于明确事实的方法和路径,承担主体应能够在已经确定的原则、标准和技术规范的指导下,具体的确定任务、任务由谁来承担,哪些由适合的工作人员,哪些由计算机,哪些由特定专家完成,要确定信息的来源和验证的方法,要确定如何得到信息,要明确计算机、软件、网络和存储资源如何获得,要确定时间表和阶段目标,确定检查或评价的实施办法。

其次是信息获取能力。构建显性信息结构,必须系统、全面获取两类信息,一是与该信息结构所表述的客观存在相关的认知信息,二是与该客观存在相关的没有包含在认知信息中的其他记录信息。信息获取能力包含支付能力,在下边讨论。除开支付能力,信息获取能力主要是发现恰当信息的能力、破解保护性障碍的能力、建立恰当信道的能力。发现能力是指找到与信息结构显性相关的认知成果信息和此外的记录态信息在哪儿的能力。对于已经发表或属于可以公开的认知成果和其他记录态信息的获取是基本的,承担主体应该建立正常的渠道,获取历史的和当前的。获取难度大的是没有明确公开与否,或原则上不公开的相关信息如何获取。要研究各种实际场景下的获取方法,如若不能全部获取,是否能有一种折中的方式,使之不影响显性信息结构的形成。这里包含三个问题,应该区别解决。一是能否知道没有公开的信息在哪里,是什么,明确与已经公开的信息之间是什么关系。对于显性信息结构本身,不需要重复的信息,最后的成果应该是最简结构,这样的判断和决策是承担主体中专家的能力。二是通过沟通和协商,获取这样的信息,这是最有效的办法,当然,也需要加以区分,不要获取对信息结构完备没有贡献的信息。三是无论以什么方式都不能获取,对信息结构完备又是不可缺少的信息,能否有办法既不违背信息不公开的客观事实,又能在一定程度上

达到信息结构完善的目的。原则上可以采用协同或委托信息拥有者（含个人及机构）不提供原始信息，只将相关部分结构化，添加到信息结构中，尽管这种方式可能没有达到应有的详细程度或留下一定的缺口，但提高了信息结构完备度。

第三是信息处理能力。信息结构显性化需要一定数量及复杂程度不等的信息处理能力。基本的信息处理能力是获取过程的传输，获取信息及显性信息结构的存储，辅助信息结构形成的处理。这里的信息处理与一般信息处理没有特别的不同，需要注意的有三点。一是对于一些信息结构，最终形成的信息结构的信息量很大，如关于一个地域的空间地理信息结构，从空间到时间，从高程到周边，从基础地理信息到各种经济、社会、生活、文化、政治、军事信息的叠加，往往一个信息点的信息描述数以万计，即使去除冗余，也是一个巨大的数量。二是对于具体显性信息结构，具有不同的结构特征，要根据这些特征，设计存取结构和算法，实现有效的存储和存取优化。三是辅助信息结构显性化的处理，既要充分利用已有的自动分词、标引、索引、主题、文摘等成果，处理原始的信息，更要在架构过程中，不断使用计算机替代人的工作，直到基本由计算机完成的高级阶段，从今天的智能技术发展趋势看，这是完全可以做到的。

第四是支付能力。支付能力是必要条件，不可或缺。在前面已经讨论了四种不同的支付模式，不再重复，只是强调，承担主体必须妥善落实费用，不仅是起始阶段，还要包括长期的维护完善过程对经费的需求。

4.6.3 协同构建与共享

在前面的讨论中，隐含了另外一种承担的模式，那就是协同构建，成果共享。不管主体居于什么目的承担了信息结构显性的工作，也不管这个领域与这个机构自身的目标是否一致，只要显性信息结构形成了，就对整个信息域所有与此相关的主体及类似信息结构显性工作具有价值。这也是信息结构显性的基本特征。

多主体共同承担，需要在统一的原则、标准和技术规范下，有效分工，

既要满足各主体自身的需求，又要符合整体规范。因此，各主体应协调成立一个工作组，负责提出整体要求并负责检查、评估、验证。如果能够以一个主体为主，其他各方为辅，则更有利于工作的进行。

协同共建形成的成果，应事先确定参与各方如何分享，也应明确如何为其他主体服务。对于这种模式，政府应给予足够的支持。

4.7 显性信息结构完备的先验性一般假定

先验性一般假定，是指在显性信息结构工作起步时先验地设定目标、框架及路径，并在实施过程中动态调整，以达到最佳实践。而推动信息结构走向显性完备的动力，是其起步、发展并走向终点的核心要素，也是需要先验地假定的。先验是指事先，未经证实的框架和路径。

4.7.1 一般假定的构成分析

先验性一般假定是信息结构显性工作起步时做出的一系列判断，以把握该信息结构显性工作的方向、路径、切入点、阶段目标和最终目标。主要包括：预期规模、信息源分布、当前成熟度、先期基础、结构特征、过程与进化、人一机互动、符合与重现、颗粒与详尽等。图 4.10 展示了一般假定的构成和分析思路。

如图 4.10 所示，对虚圈线连接的 9 个构件的判断是形成一般假定的前提，内圈三个部分是由外圈九个构件决定的，是最终决定框架、目标和重点的要素，而右侧的框架、目标、重点则是一般性假定的结果。图中的 A 是指内圈与外圈的交互，B 是指由内圈的结论决定最终的一般假定，因此图 4.10 既解释了一般假定的构成，也解释了形成一般假定的方法。

一般假定都是针对一个具体的信息结构，或若干个信息结构的集合，对这样的一个集合，我们也称为信息结构。工作量、复杂性、路径是形成一般假定的直接要素，这三个要素确定了，一个显性信息结构的框架、目标、工

作重点就可以推断出来。九个构件与三个要素之间的关系不是一一对应的，一般而言，预期规模、信息源分布、当前成熟度、先期基础、人一机互动、颗粒度与详尽度等构件与工作量有密切关系；结构特征、过程与进化、人一机互动、符合与重现、颗粒与详尽、当前成熟度、先期基础等与复杂性关系紧密；路径则与结构特征、过程与进化、成熟度、先期基础、人一机互动、颗粒与详尽等构件呈正相关。如果从更加宽泛的角度看，九个构件与三个要素都存在关联。

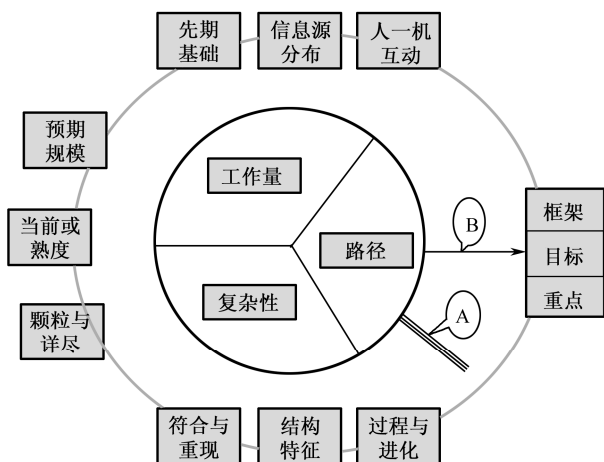


图 4.10 一般假定的结构与分析

4.7.2 预期工作量

所谓预期工作量是对一个信息结构从起步到基本完备或阶段性完备、阶段性目标所需要的工作量，即多少时间、多少人、什么样的人。一般地，还要通过工作量，最终确定工作量应该考虑复杂性和路径，提出总概算及阶段性预算。下面从规模、信息源分布、当前成熟度、先期基础、人一机互动、颗粒与详尽等方面讨论工作量的分析方法。

规模是指一个信息结构包含的信息点数和比特数，即 mit 数量和 bit 数量，

分别表示语义的量和载体与/或外壳的量。这两个量是规模计量的重要指标，相辅相成，缺一不可。对一个信息结构而言，估计规模的难度差别极大，有的几乎不用估计，能够拿出确切的数据，有的则十分难以估算。即使是对相同的客观存在，信息结构所反映的该客观存在的不同维度，或服务于不同目的，或设定不同的颗粒度或详尽度，规模就有巨大差别。在本书 3.4 中的三个例子，不管是静态空间图片、视觉神经系统的部分，还是遗传基因的局部功能规模都可能随信息结构的目标设定不同而不同。

信息源分布也是决定工作量的一个因素。一个信息结构显性相关的信息源相对集中，甚至在一个信息系统中的工作量，与信息源分布于整个信息域中大量的主体中，甚至在起步时还不知道相关的信息源在什么地方，相互之间没有直接关联的情况，工作量不可同日而语。

当前成熟度是指一个信息结构显性工作开始的时候，相对于目标而言的认知成熟度。认知成熟度不仅是指对该客观存在的认知达到什么水平，更是指这些认知是共识还是个别人的认知成果。在对一个客观存在的认知达成共识的过程中，实际上也对显性信息结构的形成相关的很多因素进行了梳理，如信息源的分布，各信息点之间的关系、所内含的功能如何实现等，这些工作都减少了工作量，反之则增加了工作量。

先期基础是指在一个信息结构显性工作开始时，与该信息结构显性工作直接相关的事项，主要体现在两个方面，一是关于信息结构显性的原则、标准和技术规范是否存在及其成熟度、可用性；二是与该信息结构直接相关的准显性结构是否存在及成熟度、可用性，如在专家系统中的知识表示、维基百科的专题中有没有相近的内容，甚至一些经典的教科书都可以成为先期基础。例如，本书第 3 章，如图 3.2 所示的街景，如果是关于图 3.2 所在空间的时间系列状态这个客观存在的显性信息结构，数千个信息点、三个月时间系列的完整描述已经存在，则第二个三个月的显性结构工作量主要就是收集信息，按框架和格式进行区分，填到相应的部分就完成了，甚至可以全部由计算机完成。相关的原则、标准、技术规范成熟度、可用性，对工作量影响甚大，特别是如果不成熟，就可能发生局部或全局的重组或重描述，可用性差就增加了制定原则、标准和技术规范的工作量。

人一机互动既是工作量的重要变量，也是路径的关键因素。显然，在信息处理速度上，人与计算机无法竞争，一旦信息结构显性工作的某些部分可以用计算机，则会大幅度减少人力的使用，从而减少了工作量。从工作量的角度看，不是是否用计算机而是能把计算机用到什么程度。从技术发展趋势看，计算机在初始阶段主要以人一机交互的方式，用于信息的收集、分类、描述信息抽取、格式化、赋予初始功能等事项，在发展中，这些事项人的参与将逐渐减少。经过一段时间的信息结构显性工作，计算机承担的事项将逐步增长，首先是关于显性信息结构初始框架构建以及评价、验证等功能的自动化；其次是越来越多的信息点、信息单元构建的功能将全部有计算机完成；第三是对成熟度评价、连接和处理的设定，信息点或信息单元的增删、组合的判断和实现逐步有人一机互动走向计算机实现；这是第二阶段。进入第三阶段，信息结构显性工作将主要由计算机完成，人承担的具体工作量逐步减少到零。尽管从全域或一个信息域的角度看完备显性信息结构的产生还需要很长的时间，但在一些局部，现在就具备实现完备的条件。

一个显性信息结构的颗粒度与详尽度要求是决定工作量的关键因素。反映相同的客观存在，一个面向具体任务的显性信息结构与一个完备显性信息结构在数量上大都存在数量级的差别，甚至存在 5 个数量级以上的差别。简单分析本书第 3 章 3.4 的三个示例，就可以得出这个结论。如图 3.2 所示的街景，在 3.4.1 中作为一般性解释，选用了 7 个信息点，而针对具体的应用可能只要一个点，如要分析危险品运输车实态，只要展开组件 2 这个信息点就行。如果是关于图 3.2 所在空间的时间系列状态这个客观存在的显性信息结构，信息点就要变成数千个，每个信息点的描述也将大幅度增加，再乘上每秒若干帧，半年或三个月的时间，两个信息结构的 mit 和 bit 总量之差就是 10 个以上数量级了。

4.7.3 复杂性

复杂性分析是构建一个显性信息结构先验性一般假定的核心主题。对于信息结构显性的复杂性，至少要说清楚三个问题：一是认知复杂性，认知基

础是否足以支撑对应客观存在的显性结构形成；二是计算复杂性，是指需要的计算资源和可调用的计算能力能否支持信息结构形成及以后生长、利用的处理需求；三是在前面两项都支持的前提下，不同信息结构显性的复杂度假定。

第一是认知复杂性。在定义显性信息结构完备时，已经明确，完备的标志是达到当时认知的最高水平，认知的突破不是显性工作组成部分。因此，认知复杂性演变为如何使当时的最高水平认知成果能够落实到显性结构中，成为本节的工作量和下一节语义通达要说明的问题，这里不做详细分析。

第二是计算复杂性。对于计算复杂性，有很多研究，什么问题可计算，什么问题不可计算。随着计算能力的增强和对问题边界与分解的深入研究，一些原来认为不可计算的变成可计算的，一些新的不可计算的问题又出现了。对于显性信息结构来说不存在不可计算的问题，因为所有的计算都是可分解的，而不可判定的问题是由人一机交互实现，并通过容错原则兜底。但对许多信息结构显性工作而言，存在计算量十分巨大的问题。计算复杂性就简化为需要多少计算能力和能否形成这样的计算能力，这一问题就演变为支付能力能否满足需求，应该由机构来回答。

第三是不同的信息结构显性工作复杂程度的先验假定。这个问题不同于计算复杂性，这是人一机协同下的复杂性。复杂性先验假定有助于恰当地确定目标，恰当地组织队伍，恰当地明确路径。以下七个方面与结构显性工作复杂性存在密切的关系，需要逐一进行分析。

一是结构特征。结构特征对复杂性的影响主要是信息结构构件之间的耦合强度和逻辑链环的长度，耦合强度越大、逻辑链环越长，对信息点和信息单元的划分、描述的规范和功能及功能间的规范和协同的要求就提升，复杂性也随之上升。本书第3章3.4.1、3.4.2、3.4.3分别例解的街景、神经系统视觉功能、启动子及转录起始过程三个信息结构的结构特征就有重大差别，而其复杂性也大相径庭。3.4.1的街景信息点之间，信息点内的描述单元之间耦合程度低，点之间及描述单元之间的逻辑关系链环短，所以由结构特征决定的复杂度最低。而3.4.2的神经系统视觉功能和3.4.3的启动子转录起始过程则相反，各信息点及信息点描述单元间的耦合强度大，全过程逻辑链环环相

扣，因此结构特征的复杂性高。

二是过程与进化。这里，过程是指该信息结构显性预期到达完备的发展阶段和时间，进化是指这个过程随着认知水平的提升和需求发展而进化的规律。这个主题将在 4.7.4 重点讨论，但它对结构显性的复杂性也存在不可小视的影响，主要反映在认知复杂性对过程和进化的影响上。

三是人一机互动。人一机互动直接影响工作量，间接影响复杂性。工作量本身就是复杂性的一个构成部分。同时，有的信息结构实现有效的人—机互动，也十分复杂。

四是重现度和符合度。信息结构显性与知识表示最大的不同在于重现度与符合度的验证，而且验证的对象是对应客观存在的对应属性与功能。尽管对该客观存在对应属性与功能的认知不属于信息结构显性工作，但符合于重现的要求，增加了工作的复杂性。符合是对已有认知及记录信息的一致性验证，重现是对显性结构所描述的功能及属性，可以在确定的条件下重现与对应客观存在一致的过程。这两个要求保证了显性结构的正确性，但增加了工作量和复杂性。在 4.9 将专门对此进行讨论。

五是颗粒度与详尽度。不同的颗粒度和详尽度不仅直接影响工作量，也对复杂性产生了直接和间接的影响。直接影响来源于有些层次颗粒度和详尽度对连接、功能和描述等显性结构构件的确立和描述难度大，如 3.4.2 视觉功能若细化到构成分子的化学功能与视细胞的化学、物理功能之间的关系，就不仅是描述量大，而且对整个信息结构的功能系统的构成和描述产生重大影响，直接加大了复杂性。间接的影响主要是工作量的不同，颗粒度和详尽度是工作量变化的主要因素。

六是当前成熟度和起步时的先期基础。起步时成熟度越高、基础越好，复杂性越低，这是不用解释的。它不仅降低了信息结构显性具体工作的复杂性，还降低了工作的计划、规划、原则和标准规范制定等基础性工作的复杂性。

4.7.4 实施路径

如何确定适合一个信息结构的实施路径是本节的目的所在，一个好的实施路径是走向最佳实践的基础。实施路径有三个关键环节，一是总目标或远期目标，二是发展台阶或阶段目标，三是起点或切入点。先验性假定的 9 个要素均与实施路径的确定有重大关系，还有一些先验性假定之外的因素，但同样影响实施路径的，例如，支付能力，即预算；近期目标，如与某个应用直接关联的结构显性。工作量、复杂性、目标、阶段性、支付能力都是影响路径的主要变量，下面逐一分析。

首先是工作量和复杂性决定了路径的确定，这是前面两小节讨论的内容，不再重复。从确定路径看，工作量和复杂性是间接因素，这两个变量通过目的设定、阶段划分等体现到路径规划中。

其次是目标和目的。在信息结构显性工作中，目标与目的并不经常相同。这里目的是指信息结构显性服务什么需求，目标是指相对于客观存在的显性信息结构完备。换言之，信息结构显性始终朝着两个方向前进：满足具体需求、实现显性结构完备。这两个方向长期看一致，但在具体的发展阶段或具体工作时，并不总是一致。一般的表现形态为满足具体需求的在颗粒度、详尽度与结构完备的要求不一致，有时还会出现边界的不一，具体任务的需要可能超越原定一个信息结构的边界。反映到路径确定上，原则就是兼顾，既要把阶段目标重点放在满足具体需求，又要把关于这一具体任务显性结构完备要求在相应的标准和技术规范中明确。

最后是阶段性。区分发展阶段，确定阶段目标是实施路径的关键环节。前面已经讨论过，不同的显性信息结构完备，实现总目标的时间、发展阶段各不相同，个体之间差异极大，如何针对具体的信息结构，恰当地划分阶段，确定阶段目标是一项重要的任务。除开信息结构自身的结构特征、复杂性、工作量等要素外，区分发展阶段，支付能力与资金的来源是重要的因素。支付能力决定了发展的速度，也在很大程度上决定了阶段的划分。而工作的起点及一个阶段工作的重点，阶段性目标的确定，主要依据资金来源的性质。如果资金是源于一般公益目标的，如一个国家或地区对一类信息的细粒度的

结构化,如英国、美国的科学数据结构化项目,人类基因组对遗传信息的测序等,起点就是一个信息结构显性完备逻辑起点。如果资金源于一项事务,诸如一个研究项目对信息收集整理,为提高研究效率而进行的细颗粒结构化;一个提供互联网信息服务的企业为有效提供服务和信息处理、存储、管理的有效性等目的,对收集信息所做的结构化工作;一条自动化生产线对相关信息满足自动控制要求的结构化;一个智能工厂从设计、制造、供应链、物流、售后服务、管理等全过程信息的结构化工作;等等。

4.7.5 信息结构显性完备的动力来源假定

动力是先验假定的一个重要方面,如果没有对动力的恰当判断,起点、目标、台阶都没有基础。作为动力的表现是支付能力和资金来源,有资金是动力的具体体现,资金的数额,体现了动力的大小。

是什么原因使国家、企业或个人愿意为信息结构显性买单呢?显然,国家是为加快社会进步、经济发展、国家安全等治理的整体需求出发,应对记录信息增长和有效利用之间的矛盾而采取的措施,从图书馆、文献中心、共享数据库到本世纪初美、英等国家的科学信息基础设施,都是源自这样的动力。传统企业的动力显然是为了更好地实现企业的发展战略和具体目标,在工业自动化、智能化之前,主要是为了研发和管理,收集信息并有针对性地结构化,当进入全面自动化,深度智能化的时期,则对自动化、智能化相关各方面的信息加强了收集和结构化。而对互联网企业而言,信息就是核心竞争力之一,更系统、全面、及时获得所提供服务相关的信息,实现与服务需求相一致的结构化,是企业的生存之道。个人的作用体现在三个方面,一是个人收集的信息的结构化,从藏书楼到个人数据库(音乐的、视频的、图像的、文字的)数千年来一直发挥着一定的补充作用;二是为公共领域或企业的信息结构化买单,为互联网服务付费就是例子;三是参与到感兴趣的信息结构显性工作中。前面两类还会继续,而第三类则由于互联网普及个体能力的提升,如果担任信息结构显性的机构采用开放平台,个体发挥的作用将日益增加,而且会改变路径。

动力来源分析不能止步于此。在直接的经济或社会效益的链环这只看得见的手之外，还有什么利益之外的看不见的手驱动信息结构的显性化，而且一直走到全域完备的终点？这只手是隐性存在的。首先是人类进化的规律。从晚期智人取代尼尔德特人的主要原因是信息表达能力的进化这个例子，决定了人类对记录信息、利用信息的追求植根于基因之中。其次是信息利用的规律。这种基因必然导致记录信息持续增长，这种增长又导致必须在一定的结构化层次上，信息才能有效利用，才能为人类本身即人类社会进步所用，说明信息结构显性是由人类对信息能力的追求带来的必然要求，而且这条路一直通向全域完备显性，信息结构显性和完备是信息发展这个链环中的一个基本规律，不可违背，也不会违背，有哪个国家、地区或领域违背了，一定会由规律把它拽回来，否则就不是规律了。第三是信息发展的规律。为什么从自在信息经由生命的自有信息发展到脱离生命的记录信息，记录信息再次发展到不经由生命的在认知意义上的自我发展，漫长的几十亿年时间里，看似无关的过程，构成了一条必然的发展链环，决定这个链环的就是信息的发展规律。

上面的分析，说明了信息结构显性完备，是由最基本的规律决定的，具体的显性工作基于特定的利益，这是一般意义上的动力机制，而作为宇宙普遍存在的信息的发展，则是由更高层面的规律所决定的。

4.8 信息结构显性的方法和工具

正如前几节的分析，信息结构显性工作量大，缺乏完整的成功实践，方法论的研究和工具的生成和完善变得更加重要。本节讨论方法论和工具的发展和演进。

4.8.1 方法论的一般讨论

信息结构显性和完备不同于过去已经存在的信息系统或软件系统的开

发，也不同于数据库或大数据的组织、管理和利用，需要有适合其特征的方法。

一般而言，信息结构显性的起步与起步后的发展与完善具有重大的不同，因此，方法论也如图 4.11 所示一样，由两个相互连接的部分构成。

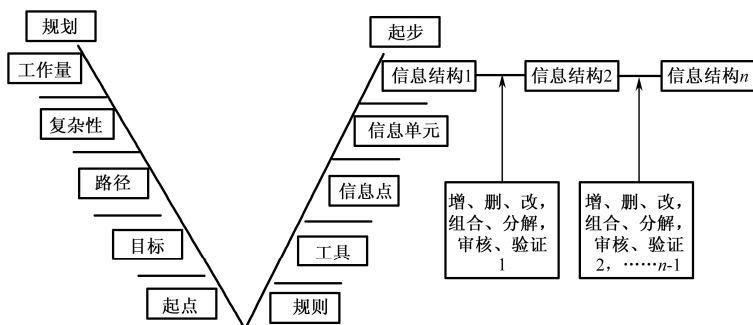


图 4.11 信息结构显性的一般方法论

图 4.11 展示的方法论由一个 V 字形构成的初始阶段及一字排开、持续滚动前行和完善的发展阶段构成。在第一个阶段，V 字形的左边是前一节讨论的内容，是决定起步之前的所有工作。由于迄今为止，没有一个完整的信息结构显性实例，所以这一侧的工作称之为先验性假定。也就是说，规划是建立在部分相似工作基础上的推论，需要用推论的方法，而不是归纳的方法来制定。经过若干年之后，有一定数量的信息结构显性达到了当时认知水平的基本完备，则会从现在的推论为主转向归纳为主，但依然存在一定先验假定的需要。

V 字形右边一侧是在规划结束，确定了起点之后到最初一个显性信息结构成型的过程。首先是制定规则，就是该信息结构显性过程要遵循的原则、标准和技术规范。要按照显性信息结构完备的五个测度、十个一般性原则、六个标准和四项技术规范来具体制定本信息结构显性工作要遵照执行的原则、标准、技术规范。目的是保证工作的有效性、规范性、可持续性和结构进化要求。其次是准备工具。信息结构显性的特性决定了需要很多工具，这是 4.8.2 讨论的内容。由于还没有形成直接服务信息结构显性的工具，需要通

过借用类似工具进行二次开发，有的则需要从头开发。第三、第四、第五是从信息点开始的信息结构构建工作。

第二部分是一个动态的滚动发展和完善过程。每经过一个工作阶段，信息结构通过获取、分解、组合，形成了一个新的实态，这个实态要经过审核和验证，来确认已有的成果，并评价方法、工具和路径，确定下一阶段的目标，修订或保持原则、标准、技术规范 and 工具。一次次的发展和评价、审核、验证，逐步向基本完备趋近。

4.8.2 工具的形成

信息结构显性工作 need 一组工具的支持，包括形式化规范工具、属性描述工具、功能描述工具、结构解析工具和验证审核工具，其构成和关系如图 4.12 所示，各个工具的作用、特征、来源、相互关系在下面分别介绍。

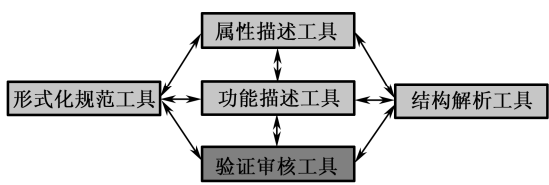


图 4.12 信息结构显性的工具体系

图中功能描述工具框中的虚线表示属性描述工具与审核验证工具之间和形式化工具与结构解析工具之间的连接通道，也就是说工具之间是全双向网状联通的。

4.8.2.1 形式化工具

形式化工具是对信息结构显性过程的格式、结构和规范要求通过工具来遵循和实现。至少应该包括两个系列八个主要功能。

第一个系列是关于本项工作已定标准和技术规范的形式化解析，形成格

式化的工具，规范整个信息结构显性过程各个步骤，使之符合标准和技术规范的要求。这类工具应该在标准和技术规范制订时同步开发。要注重把标准、规范中的一些软性约束格式化，把所有要求与相应的流程挂钩，保证在实施过程中落实。

第二个系列是关系信息结构各构件的形式化。主要包括不同类型信息点的组成部分及相应的格式要求，不同类型信息单元的组成及格式要求，一个信息结构的组成及格式要求。形成如图 4.13 所示的体系性格式规范。

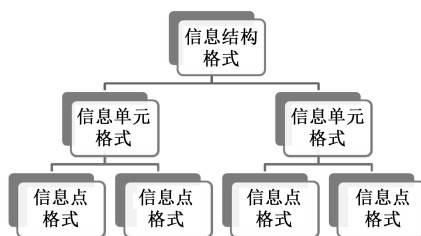


图 4.13 一个信息结构的体系性格式规范工具体系

其中，工具对信息点的格式规范主要包括但不限于：对所描述类型做出规范，所有的描述都可以归到一个或几个类型中；对所有的描述类型的描述内容和格式提出确定的要求，规定描述中不同内容的排列次序；对描述方法提出规定，对数值、文字、图片、音频和视频等描述客体的规范；对描述如何与感知、连接和处理要求的一致和优化做出规定并提出实现的形式性办法；对地址、感知、连接、处理的标识做出精细的规定等。

对信息单元的格式规范主要包括但不限于：依据本单元信息点间关系及各描述体的语义关系确定的体系和物理布局，如文本与图片（见 3.4 个例示）、音视频等记录信息，主描述体与辅助描述体区分的标准和的格式；本单元各功能的总体描述及其格式要求；根据本单元功能的特征确定对感知、连接、处理的基本描述格式，如感知、连接、处理的主要类型的格式，功能描述与重现的结构描述格式规范；等等。

对信息结构的格式规范主要包括但不限于：对本信息结构各构件的命名规范；对本信息结构各类属性的分类规范及描述规范；对本单元的公共信息

单元的定义和格式；本信息结构与外部连接的分类及格式规范；等等。

4.8.2.2 属性描述工具

属性描述是信息结构显性过程中工作量最大的部分，工具越贴近工作实际，支持的功能越多，工作量的减少也越多，而且质量也会提升。属性描述需要工具支持的内容众多，主要有：信息获取的工具，信息摘取或描述体形成的工具，功能属性和事实属性的区分和描述规范的工具，信息点、信息单元、信息结构 3 个层次描述的区分和实施工具等。

信息获取的工具是指从信息源获取信息结构显性所需信息的工具。结构显性所需信息的来源主要有三类，一是互联网上的开放信息，二是特定对象的记录信息，三是特定目的的记录信息。其中第二类相对于信息结构反映的客观存在，第三类相对于具体任务的目标，有时候两者之间会重合，但大部分场景下不一致。结构显性所需信息从记录存在状态看，可分成已存在和与结构显性同步感知的信息。前者容易理解，大部分结构显性使用已经存在的信息，但也有一些会与记录信息的产生同步或准同步。例如一个以实时路况信息为起点的信息结构显性，3.4.1 的例子就可能发生这一状态。不同的信息来源和信息存在状态，需要不同的工具。互联网开放信息源主要是抓取工具，这类工具已经大量存在，选取类似的工具加以适用性改造即可。特定对象的记录信息获取工具更加简单，是一个能从对方信息库中直接抽取的工具，这种工具基本上已经存在，也只要适用性改造就行。特定目的的信息源视实际情况而定，可能是前两种之一，也可能两种兼而有之，工具的形成亦同。与记录信息生成同步或准同步的信息源，获取工具就是从中选取需要的部分，不管是什么类型的实时信息，只要能用到相应的系统中，就能构造相应的工具。

功能属性和事实属性的区分和描述规范的工具是指一个信息点、一个信息单元或一个信息结构的描述中，按照一定准则区分功能属性和事实属性的工具，以及这种区分在描述中如何实现的规范性工具。在一个信息结构中，理论上存在两种不同的信息单元、信息点或描述，即一类主要将对应客观存在的事实结构化描述清楚，另一类是将对应客观存在的功能描述结构化清楚

并能按这一描述重现。但是在表象上，两类都是一种事实性描述，只是目的、格式、作用不同。因此要通过一种工具，并与其他工具协同，先行做出区分，并按不同的类型给出描述的格式和要求。

信息摘取及描述体生成的工具是最重要的属性描述工具，因为生成描述体所占工作量最大。该工具是指从已经获取的相关记录信息中摘取构建显性信息结构有用的部分，并按照相应的规则和协同其他工具，生成描述体的工具。一个信息结构，甚至一个信息点，不同描述体自动生成的难度差异巨大。有的描述体内容很容易从相关信息中自动摘取出来，再按照规定的框架、格式和形式加工处理，工作就完成了，这样的例子不少，如时间—空间特征明显，描述体沿给定时间周期重复的类型，不同信息点同类功能的描述等。有的描述体语义和功能关系十分复杂或信息源本身特征不利于摘取工具的使用，对描述体生成工具提出了复杂性挑战。不同类型记录信息的需要不同的摘取工具，或利用综合工具中的相应功能。3.5介绍了一系列从不同信息源摘取结构显性所需信息的模型或算法，摘取工具需要利用所有相关的成果。描述体生成工具的发展，可参考自动文摘等类似的成果，但应针对一个信息结构的特征来设计并开发适用的描述体生成工具。

信息点、信息单元、信息结构三个层次描述的区分和实施工具。它有两项主要功能，一是根据形式化工具给出的框架，将一个具体的描述体放在结构的什么层次，二是按照适度冗余的原则，确定在什么层次、按什么规则保留什么样的冗余，主要针对事实性描述。一个信息结构对一个描述体放在什么位置，需要全局优化，主要决定因素是存储和计算的复杂性。放在信息点、信息单元还是信息结构的整体描述中的决策需要工具支持，工具计算不同位置的存储和计算复杂性，然后做出判断。信息结构要完成全面的重现功能，信息点、信息单元要完成相应的重现功能，如何科学地分配相应的功能，优化重现过程，是本工具的主要任务之一。同样的事实或功能描述，在不同的信息点和信息单元都需要，采用冗余描述还是指针引导，依赖于计算模式和计算量。实际上，这个工具也是相关功能实现工具的辅助工具。

4.8.2.3 功能描述工具

功能描述工具主要针对信息结构相对于客观存在的功能重现。在信息结构显性的每一个阶段都要做到已经显性部分所反映的客观存在的相应功能能够真实重现。一个信息结构关于对应客观存在的功能部分能够重现，必然需要一套有效的工具支持，主要包括：感知功能描述工具、连接功能描述工具、处理功能描述工具、综合功能描述工具。

感知功能描述工具是指在信息结构显性过程中对感知功能描述和重现提供支持的工具。该工具的主要功能是：本信息结构感知功能的全部类型及其实现方式；各类感知功能的在一个信息结构中描述的整体布局、功能划分和实现；各感知功能对连接的需求及其实现方式；各感知功能对处理的需求及实现方式；确认一个可能是新的感知功能的真实性，如果是真，则判断其类型；在信息点、信息单元、信息结构新增描述点之后，判断是否与感知相关，如果相关，判断是否需要建立感知连接，判断是属于已有的感知类型还是需要建立新类型；等等。

连接功能描述工具是指在信息结构显性过程中对连接功能描述和重现提供支持的工具。连接功能是显性信息结构中最重要功能，也是功能重现的基础。该工具的主要功能是：本信息结构连接功能的全部类型及其实现方式；各类连接功能在一个信息结构中如何描述的整体布局、功能划分及其实现；各连接功能对感知的需求及其实现方式；各连接功能对处理的需求及实现方式；确认一个可能是新的连接功能的真实性，如果是真，则判断其类型；在信息点、信息单元、信息结构新增描述点之后，判断是否与连接相关，如果相关，判断是否需要建立新的连接，判断是属于已有的连接类型还是需要建立新类型；等等。确定连接的类型要细致周到，至少包括下面几个大类：稳定的永久连接；基本稳定，但存在变化可能的连接；临时连接；随时可能发生的任意连接；语义的连接；功能的连接；描述的连接；结构层级的连接；等等。

处理功能描述工具是指在信息结构显性过程中对处理功能描述和重现提供支持的工具。处理功能在显性信息结构中十分重要的功能，是功能重现的主要构件。该工具的主要功能是：本信息结构处理功能的全部类型及其实现

方式；各类处理功能在一个信息结构中如何描述的整体布局、功能划分及其实现；各处理功能对感知的需求及其实现方式；各处理功能对连接的需求及实现方式；确认一个可能是新的处理功能的真实性，如果是真，则判断其类型；在信息点、信息单元、信息结构新增描述点之后，判断是否与治疗相关，如果相关，判断是否需要部署新的处理，判断是属于已有的处理类型还是需要创建新类型；处理功能在信息点、信息单元、信息结构整体的部署进行优化的工具；等等。

综合功能描述工具是指一个信息点、信息单元、信息结构除上述三类功能性工具之外的工具，主要功能是协调感知、连接和处理功能，实现这三种功能的整体布局优化，从达到重现度要求的目标出发，描述需要补充的其他功能。说到底，这一工具就是以本信息结构所反映的客观存在的功能重现为出发点，从信息点、信息单元一直到整个信息结构，全面梳理感知、连接、处理三大功能工具，为信息结构正确实现重现功能，协调分配相应功能，优化重现过程。

4.8.2.4 结构解析工具

结构解析工具有两类，即语义解析工具和结构解析工具。语义解析工具的主要功能是维护信息结构的语义结构图，将信息结构所有描述所涉及的全部概念与语义结构图一一对应。这个工具本质上就是信息结构的概念模型，规范信息结构新概念的使用，并为此建立语义关系。按照适度冗余的原则，在信息结构中容许同一概念，同一描述在不同的位置重复出现，平衡的关键点是计算和存储的优化，而不是结构的简洁。同样，在信息结构中容许同义词、近义词的交叉使用，判断的准则不是为了易处理而规范为一个标准的词，通过语义解析工具中的关系来实现连接，而是在一个具体的信息点或信息单元中，何种表述更适合、更清晰就使用这个表述，而概念间关系则同样通过语义解析工具来完成。

结构解析工具可以称之为结构的结构，通过表或图的方式，将已经建立的显性信息结构简洁又完整地表述出来，其主要功能是为信息结构显性和完备过程中的各项操作做出判断。

一个信息结构在显性及完备过程中，主要的操作是各个组件的增、删、改，组合与分解，组件间或结构间的映射。触发操作的是接收到与信息结构显性工作相关的新信息，或在审核验证中发现的问题。结构解析工具在做判断时应与语义解析结构等其他工具衔接，以做出一致的判断和决策。结构解析工具是在触发后确定实施什么操作，而不是操作本身，操作是由结构内含的处理功能实现，这就是为什么各项功能要从信息点开始一路上行的原因。

增删改操作比较明确，也比较容易实现，在很多领域都有相关论述，这里不加讨论。信息结构内外不同层级、不同组件的组合、分解和映射是结构解析工具的重要内容，其对象又有特殊性，是形成该工具需要重点研究的主题。组合和分解是一种特殊的增删改，它不是简单地增加或删除，而是信息结构的一个构件或一组构建的集体操作，组合将不同的两个或两组合为一个，分解则是将一个或一组拆分为一个、一组或几个、几组。在实际过程中，还会发生组合、分解的连锁，一个过程有若干次组合或分解构成，而且在这个过程中，对象集合也可能发生变化。

结构映射是验证审核过程的功能性需求，更是显性结构增长的重要方式，结构映射在不同的领域有不同的含义，定义 4.4 确定了本书的工作定义。

定义 4.4: 结构映射是指同一信息结构或不同信息结构中两个完全对等的组件集合进行比对，并确定差异的过程。定义规定了结构映射的条件，必须是两个完全对等的结构集合，是否完全对等，依据是已经存在的结构解析工具给出的“结构的结构”。结构映射的结果是两个集合比对后确定的差异。确定的差异用于验证或审核就会发现存在的问题，更重要的是用于比对的两个集合自身的结构完善，差异就是完善的方向，甚至可以直接进入结构完善的操作。

结构映射需要同时使用结构解析和语义解析两个工具。构成映射的两个集合及比对算法的设定，主要依据结构解析工具，而确定比对后的差异，主要依据语义解析工具。

4.8.2.5 验证审核工具

验证和审核在工作对象、内容等方面看，不能从语义上做出明显的区分，

在本书中约定，验证是指信息结构显性的主体内部对工作成果的评价，而审核是指来自外部的评价。

验证或审核工具的功能就是针对作为验证或审核对象的显性信息结构展开评价，给出合规性、正确性、完备性等方面的结论，指出不足之处。合规性就是与设定的十项原则、六项标准、四项技术规范进行比较，看是否一致。正确性就是对各信息点、信息单元及整个信息结构从形式到语义，从属性到功能，从描述到重现，全方位进行检查，得出哪些是正确的、哪些是不正确的结论。完备性是把该信息结构已经实现颗粒度、详尽度、符合度、重现度等主要指标，与该信息结构所反映的客观存在的当前认知水平和记录信息已经达到的高度和宽度进行比较，得出相对完备度的结论。

验证或审核工具应建立在本节所述各类工具的基础上，但要根据相应的任务，保证评价过程中的中立性而进行二次开发。

4.8.3 开放、开源和交互平台

信息结构的显性和完备具有工作量巨大、受益面广泛、有能力参与人数众多的特点。针对这样的特点，通过互联网平台，采取有管理的开放和开源路线，是使信息结构显性和完备能够跟上记录信息增长步伐和应用需求的唯一可行办法。

4.8.3.1 开放、开源的一般原则

互联网为实现开放和开源打开了大门。大门打开了，如何使数以十亿的可能人群有效参与，需要建立可操作的开放、开源原则。

开放的原则是，只要是互联网的用户，有意愿、经验证，能为众多开放于网上的信息结构显性任务做出贡献的人都可以进入这一平台。这个原则无约束、全开放，不问身份、不问学历、不问国籍、不问专业、不问目的、不问能做什么工作和多少工作，只要有能力、有意愿就行。验证是指能力的验证，而能力的验证实际上就是通过做网上开放的任务来实现。这里的参与不仅是指对已经发布的任务的参与，承担其中的一个部分，也包括提出一个或

几个新的信息结构显性任务，并承担主体的责任。

开源的原则是，只要是在开放的平台上实现的或准备在开发你给的平台上实现的结构显性及相关资源，奉行有管理的全开源原则。全开源包含所有与信息结构显性相关的资源，原则、标准、技术规范、工具，已经完成的显性信息结构、已有并可用于信息结构完善的记录信息集合，用户侧之外的可用处理能力等。奉行全开源原则一是与全开放原则匹配，没有全开源，全开放就不能落实；二是充分利用，实现社会效益最大化。当然，开源原则受利益机制的约束，如果是为特定的商业利益，用商业投资进行的信息结构显性成果是否开源，由投资主体依据利益诉求和平衡，做出决策。

4.8.3.2 交互平台及其管理和利益处置

全开放和全开源的原则必须建立在有效管理的基础上，没有全平台同一的有效管理，开放和开源必然导致与预期相反的结果。管理有三个主要任务，一是对结果的验证和审核，二是对参与者的激励，三是对所形成成果的利益处置。

在开放平台上参与的志愿者不是专业人员，也不是专职工作，一般不关注，也难以评价结果的质量。所以管理的第一项任务就是保证所有在开放平台上形成的显性结构组件的质量进行验证和审核。这里同时用了验证和审核两个词，主要表达任何平台上形成的成果都要经过多次验证或审核，而不仅是上节提到的内部和外部的关系。验证或审核主要通过已有的工具，必要时，也可以用开放、开源的办法进行验证，但经过这种方式验证的结果，还要经过管理者的验证和确认。尽最大可能保证结果的质量是开放平台管理的首要任务。在结果验证或审核管理中，也包括了参与者新提出的信息结构是否采纳的决策，要根据已有的成果和正在进行的信息结构，做出是否同意增加该信息结构，还是纳入已有或已经在进行中的信息结构中。

管理者必须提出如何激励参与者的办法。除了极少数具有显著商业利益的结构显性任务，主体会根据工作量给予报酬，绝大部分结构显性任务是基础性的社会公益活动，因此没有报酬。没有报酬的志愿行为如何持久是激励机制的一大难题，但又必须解决，以保证本质上要长期发展的信息结构显性

能持续不断，直至达到完备。平台必须创造一种以成果为基础的荣誉性激励机制，使参与的人根据贡献大小，能够随着其参与的信息结构存在而得到长期的荣誉激励。

管理者必须对所形成成果的利益提出处置的原则和具体办法尽管绝大部分显性信息结构没有直接商业利益，其主要贡献是提高全社会发展的质量和速度。但不可否认，某些信息结构显性成果，或其中的部分对一些商业行为具有使用价值，而开源就意味着有些企业或个人可以通过开源原则获取利益。因此要明确商业用途必须购买的管理原则，要通过管理工具追踪所有开源显性信息结构的使用，确认其是否具有商业利益，并有手段对不知福的商家或个人惩罚或终止其使用。而对受益，也要有以长远发展和参与者贡献相结合的分配方式，分配方式也要开放、开源，资金从哪里来，如何分配在全平台公开，用全公开实现公正、公平。

4.8.3.3 从串行到并行的历史性转折

互联网改变了很多领域的发展模式，从封闭到开放，从私有到开源是历史性大趋势，对于信息结构显性而言，更重要的是形成了从串行发展到并行发展的平台。

互联网环境下信息结构显性的全开放、全开源既有与一般领域相同的地方，也有不同的地方。相同的是都是开放开源，只是更加彻底、全面，不同的是信息的增长以互联网平台上的开放开源为契机，历史性地从串行或小规模、小范围并行走向大规模、全面并行，实现了发展方式的质变和发展速度的质变。这双重质变既是信息增长的拐点，更引发了经济和社会发展的全面转型。

从历史看，从生命感知信息，引发对环境变化的反应和遗传基因的突变和选择，到语言文字的产生走过了30~40亿年历程；从具有概念性质的语言文字产生到记录信息生产、存储、处理、利用的数字化、网络化又走过了约10万年；可是在数字化、网络化环境下，记录信息的增长呈现出不可思议的速度，最近两年新增的信息是等于或接近于除这两年外人类记录信息的总和。这一速度出现根本的原因是几十亿用户成为信息从自在态或自用态向记录态

信息转换的主体，发展模式从串行发展到并行，以亿为单位量级的并行，产生了排山倒海的力量。

而记录信息的增长改变了过去只有少数人拥有信息，信息不对称成为许多学科研究的领域，也成为不平等、不公正的一个源泉，信息不对称条件下的博弈成为一门学科。数字化、网络化彻底颠倒了这一关系，在信息的汪洋大海中，存在着人们解决当前问题十分有用的信息，但不能从中吸取足够的所需信息。信息“爆炸”这个词，说明了人类对快速增长的记录信息缺乏足够的去及时、有效地利用，任何个体的生理能力，在认知提升（科学研究）和问题解决两个主要领域，都不可能在一个应该把握的领域、利用应该利用的记录信息，推动发展和进步。

改变这种状态，加快信息结构显性，在一个个主题域、信息域、直至全域形成完备的显性信息结构成为关系到人类认知能力发展，关系到经济社会发展速度和质量的关键环节。幸运的是，互联网为代表的信息技术发展为解决这个问题展示了曙光，通过互联网平台，吸引全部（至少大部分）互联网用户参与显性信息结构的构建，用全域（或信息域、主题域）最简完备显性信息结构来应对这个挑战。如果我们将坐标系做一个旋转，以数十亿人的并行工作和数十亿年的串行发展放在同一个坐标下，就会发现只有这种大规模并行才能使信息增长的两个方面，记录信息和信息结构显性，达到平衡，从而使人类有可能在信息“爆炸”时代跟上它的步伐，而不被增长的记录信息所左右。

4.8.4 工具的演进

信息结构显性的工具在发展中将沿着三条路径演进，一是在一项任务中的串行演进，二是平台环境下工具的细分，三是基于平台的并行发展。如同对结构显性的开放和开源一样，对工具的演进也需要必要的管理和控制。

4.8.4.1 基于任务的串行演进

4.8.2 中对各类工具的描述原则上基于一个任务，完成一项任务的过程

中，工具必然需要并实现演进，这是工具演进中最基本的模式。基于一项任务的演进本质上是串行的。随着任务的进展，工具会产生三种变化，功能增加、协同完善和错误修正。

首先是功能增加。在信息结构显性的发展中，必然会遇到一些新的功能和事实，有的是新的描述类型，需要在描述工具中增加；有的是具备或需要新的感知、连接、处理功能，这些功能如果不能通过相关工具中已有的类型处理，就需要增加新的工具；有的是在格式或形式上超越已经确定的模板，需要新模板；有的是在语义构成上超越了已有工具的语义解析范畴，需要增加新范式；有的是超越的结构解析工具的范畴，“结构的结构”不能表述这个新的信息点，结构解析工具要增加新成员。

其次是协同完善。一组工具系列相互存在很多耦合关系，当一个工具产生了增加或错误修正，其他的工具应该随之改变，因此，工具的演进也需要一个工具来协同。

第三是错误修正。错误修正是增加的反向行为，在结构显性化过程中，一定会发现原来工具中存在的错误，需要及时修正，这样的修正也要反映到其他工具中。

4.8.4.2 平台环境下的工具细分

在一个项目的发展中，必然会发生如果工具更加专门化，效率和质量会更好，演进也会更加频繁和有效。一个项目涉及的对象太少，不足以推动项目的管理者修改工具，但进入多项目并行使用的平台就不一样了，工具细分带来的好处会叠加。只要有足够的使用需求和经验，可以把工具细分到无须人的参与就能够自行决定的水平，如在 3.4.1 中，表 3.9 视杆细胞结构描述中，将工具细化到对基础部分、感知部分、连接部分、处理部分的文字描述、结构描述、功能描述和图形描述的 16 个工具，用来描述不同任务中同类细胞结构，如果有足够多的任务，这 16 个工具的效率和质量将持续提升，需要与人交互确认的内容越来越少，直至可以自动完成一个新信息到来后的结构化任务。

细分的作用是显然的，问题在于要明确谁推动、谁办理、谁验证、谁决

策。推动者应该是平台的管理者，通过提出需求和激励制度来推动工具的细分；办理者应该是平台上参与者，将与自己正在进行的任务中使用的工具细分；验证者应是各相关具体项目执行主体规定的验证者；决策者应该是平台的管理者，从起始到确认的流程和办法应有平台管理者确定。

4.8.4.3 基于平台的并行发展

这里平台是指在互联网上用于信息结构显性的工作平台，这个平台不是服务于一项任务、一个信息结构，而是在一定的范围内所有信息结构显性任务的共同平台，并行发展基于这样的平台。

并行发展是指不同的项目、不同的信息结构显性过程中相同工具的协同发展。信息结构显性工具的发展具有基于经验、试错的鲜明特征。基于经验和通过试错来发现能够改进的环节和需要修正的错误，这类事物加快演进的一般规律是加速迭代频率，加快发现可改进和需修正的问题。

假设一个信息结构显性工作一年在某个环节发现能够改进 m 点，修正错误 n 点，如果一个平台上有 10 个、100 个、1000 个类似的任务，尽管不是直接相乘的关系，但也会以数量级的速度加快，而且由于参与的人数在增加，还会提高工具完善的质量。信息结构显性内在的可分解、重复性、低进入门槛、大工作量等特征，决定了工具完善并行模式的可行性和重要性。

4.8.4.4 工具演进的管理和控制

对于一个任务的串行工具进化，显然由这个项目的主体承担管理和控制责任，而对于互联网平台上的大范围并行进化，则需要确定管理和控制的规则。规则至少应包括三个方面：动力和激励机制、工具和方法、管理和控制准则。

动力和激励机制是前提。由于经济性动力机制在一般意义上不存在，所以，动力主要是有利于参与者减少工作量，提高工作的效率和质量，根据贡献大小的全平台排名等方式。按贡献的全平台排名需要有一个委员会式的组织，负责全平台的统计和甄别。如何提高参与者从中获得的利益，关键就是贯彻好全开放、全开源的基本原则，无条件地贡献自己的改进、获得别人改进的成果，使这一原则取信于所有参与者。

工具和方法是保证改进的有效性，提升参与者这一事项中工作效率的有效途径。因此每一个工具应该有一个模块，有一个制度或流程，专门服务于广泛参与的并行改进工作，既有利于参与者提交改进的具体建议，也有利于工具的管理者确定修改与否及具体的修改操作。

管理和控制准则既需要在全平台这个范围内有个需要共同的管理和控制的准则，对各个具体工具的管理者明确要求和责任，对所有参与者提出责任、义务和权利，对申诉、争议应建立仲裁机制；还需要对具有商业利益的行为制定原则规定。

4.8.5 长期进化

记录信息快速增长是不可避免的，人类的认知能力跟不上记录信息的增长，也是不可避免的。在科学研究、复杂经济和社会系统的管理、人类生活适应“信息爆炸”带来的新环境等领域，没有最简、完备、显性信息结构作为基础，将会变得很难适应，并不可避免地出现震荡。这种局面凸显了加快信息结构显性的步伐对人类社会发展的作用。

如果我们把记录信息的增长及其信息结构的显性作为一个客观存在，信息结构显性的工具就是要把这个客观存在的对应信息客体结构化。从这个角度去分析，就可以把握长期进化的方向和路径。

方向就是结构显性最终与记录信息的增长同步。路径就是逐步将信息结构显性这一事务成为自动化的流水线。记录信息一经产生，除了集聚到需要使用的地点或系统，同时流向自动化的结构显性生产线，成为该领域相应显性信息结构完备的新原料、新需求。利益关系则通过原始记录信息的所有权和使用权来规范。

通过自动化来应对供给不能满足需求是社会发展历史总结的基本经验。能够通过自动化来实现记录信息增长与显性信息结构完备的平衡，并以显性结构为中介实现记录信息的增长与利用的平衡，是因为信息结构显性工作本质上是简单的复杂系统。

所谓简单的复杂系统是指如果将信息结构显性工作分解到一个具体的信

息点，一个具体的信息点的一个感知、描述、连接、处理，并且这种处理是在规范好用的工具的指引下完成，因此十分简单，绝大部分这类细化的具体工作并不需要高深的知识或技术基础，中学生、本科生都能完成，特别是工作量特别巨大的常识结构化，基本只要具备上网能力的人都可以参与，只有少数具有开创性、前沿性的信息结构中的部分信息点、信息单元需要这个领域的高水平人才才能担当。但是，将这样一个个简单的信息点、信息单元、信息结构作为整体，特别是将一个主题域或信息域作为结构显性的对象，这个系统又变得十分复杂。本质上可分解、可并行的信息结构显性就要依据这样的特征组织实施，研究其发生发展的规律，确定方向，实现长期演进的目标。

4.9 信息增长的规律和轨迹

本章已经讨论了信息增长的定义和发展阶段，介绍了共自在态信息到遗传信息、认知信息、记录信息的增长过程和规律，讨论了信息结构及其显性在信息增长中的作用以及信息结构如何显性，在最后，还要进一步讨论如此形态特征、参与主体不同的信息增长的一般规律，今后一个阶段的主要发展方向和对信息处理的新需求。

4.9.1 独特的增长过程和规律

本章 4.1 节中详细讨论了信息增长的阶段，从自在态信息到遗传信息，再发展到认知信息，认知信息的发展又产生了记录信息，记录信息的增长导致了显性信息结构的产生，完备的显性信息结构又为自主的信息空间的生成创造了条件。这是一个独特的发展过程，既不同于生命体，又不同于非生命的自然物质。每一个发展阶段似乎都是独立的，有着自身的发展规律，但连接在一起，又跨越了分离的阶段，在各自发展规律上又形成了统一的规律。

这个统一的规律是什么？如果没有生命的产生，自在态信息就不可能成为信息增长的起点，但是生命发生了。作为环境对生命影响的载体，自在

态信息成为遗传信息突变的触发器，也是突变的遗传信息被选择的原因，生物体从环境感知的信息成为判断优劣的依据。生命和环境交换的不仅是能量和物质，能量和物质的交换保证了生命的延续；还交换了信息，信息推动了进化。从最简单的生命到具有复杂认知功能的人类的产生，这是一个长达约40亿年的岁月。

一旦在漫长的功能进化达到了感知与神经系统能保存记忆、发出一个生物群体共同理解的声音，这些声音代表着不同的概念的时候，记录信息的产生就不可遏止地发生了，地球上几个不同地方文字各自独立产生说明了这是规律。文字的产生就是记录信息产生的标记，一经产生，发展就与经济社会的发展紧紧连接在一起，不断加速，在文字和第一个记录信息产生到现在大约8000年的历史中，世界各地不同的文字和文明殊途同归，各自演化出不同的文字，就是一个有力的证明。语言文字的产生和使用，在不同地域发展的不同的语言文字系统演化出日益复杂的概念系统，独立进化的不同语言往往具有相同或相似的概念表达类似的事物，这是今天能够对不同的语言体系进行翻译的基础，不能不说这是概念体系进化的规律；一个地区的文化历经文明的跌宕起伏，但概念系统如同植入社会的基因一样，历经近乎毁灭性的灾难，依然在后续社会中延续，如秦始皇焚书坑儒，试图将儒家文化连人带书（概念体系）葬送，然而儒家文化依然同华夏文化一并发展，历经数千年长存。

工业革命以来，记录信息的增长持续超过了经济社会发展的速度，计算机、通信网络、相片和音视频记录设备、传感器为记录信息的增长在生成、处理和存储、传输三个方面发力，互联网更成为指数式增长的催化剂。为了更好地利用记录信息，从记录信息产生的早期开始到今天，人类做出了持续的努力，藏书楼、图书馆、分类、编目、索引、数据库的产生和发展，就是跨越几千年努力的见证。但记录信息的增长不仅超越了经济社会发展速度，更超越了人的认知和利用能力，快速增长的记录信息和缓慢发展的人的信息利用能力成为主要矛盾。需要有一条路径减缓这种矛盾的进程，更需要一种方法能在记录信息急剧增长环境下，保证地球文明的演进不停滞。显然，记录信息的增长是不可遏制的，人的认知功能进化是极为缓慢的，这一规律不可改变。唯一的道路转向信息结构显性。记录信息结构显性在两个方面提供

解决方案，一是通过最简完备显性信息结构来为信息的利用提供最简捷路径，减缓矛盾；二是在信息结构显性发展到一定程度，摆脱人的干预，自动生成显性结构。这是一个重大转折的起点，自动生成显性信息结构意味着系统对记录信息的自动理解，而且是达到当时最高认知水平的理解，那么，只要对系统在增加一定的推理、演绎功能，就可以实现自行或与人类交互，但以系统为主的认知发展。走到这个转折点，一条彻底解决记录信息快速增长与人类认知能力进化缓慢这对矛盾的路径就产生了。由此，以一个个问题场景、一个个主题域、一个个信息域的简洁显性信息结构为基础，构成了全域简洁信息结构，该结构成为全部自有信息、记录信息的参照系，成为所有主体利用、发展记录信息（包括认知信息）的纲，成为自在态信息最完整的映射，从而一个独立的、持续发展的信息空间就产生了。

这种跨越认识论范畴的主体和客体，跨越不连续的物理过程、生命过程和社会发展阶段，实现如同一个物理、生命过程，一个社会发展阶段的持续增长，是信息增长独有的规律，研究并认识这个规律，是改变地球文明演进必然碰上的困境的重要任务。第6章将进一步阐述这一命题。

4.9.2 跨越转折点

信息结构增长正处于一个重大转折点上，这是信息增长的最后一个关键转折点，走过这一步，信息增长进入自我发展的新阶段。成功跨越转折点的关键环节有两个：记录信息增长的有序和信息结构显性取得全过程示范性进展。

记录信息增长的有序要解决两个问题：一是应该记录的信息没有记录，二是如何使信息的记录、存储和处理实现全局优化。不可能记录所有的自在态信息，数学工具和实践都可以证明这是做不到的，一个个具体的有限记录，不可能穷尽无限的自在态信息。记录需要有选择，不同的主体承担不同的责任。国家应该研究并列举所有应该由国家承担的信息记录职责，特别是现在不记录，会对以后的研究或管理带来缺憾的自在态信息记录，如典型性、重大社会事件，各类具有科学研究价值的自然状态的记录；机构应该承担起本机构相关的事件或状态的记录，并将各种记录动态梳理，既没有重大遗漏，

也没有过多冗余；个体也应该承担起自己的责任，记录什么，不记录什么，如何保存，如何处理要有个计划，既不要对以后的生存或延续过程留下缺憾，也应与社会处理能力和个人支付能力一致。记录、处理和存储需要成本，有的成本还很高，特别是对能源和物理空间的需求。在源头不同、快速增长的前提下，需要在国家、区域、机构和提供服务的企业层面，对处理和存储实现优化，国家还应该对如何优化、全局优化提出可用的标准和技术规范。

信息结构显性取得全过程、示范性进展是指在一个有一定规模，如一个难度适中的主题域，该域信息结构显性基本达到完备，并从确定原则，制定标准和技术规范，研制工具，确定实施主体，基于互联网的开放和开源，工具的完善、各种激励机制的确立和实施，各项验证和审核，特别是详尽度、颗粒度、符合度、重现度以及工具等方面，各环节自动化的比重等均得到该项目的验证并实现。一个主题域的全过程示范，既验证了方法的可行性，更重要的是实质上确定了结构显性的标准化流程和方法，对以后的发展在质量、速度和动力的积聚都有重要意义。

对于信息增长，尤其是显性信息结构的自动化生成，一定有人担心如果失控怎么办。对于显性信息结构增长的过程控制，特别是进入第六阶段之后的控制是十分重要的，但本书的显性结构完备是以已有认知成果和现实客观存在为基础的，没有超越已知的学习和探索，因此不存在失控的问题。至于在显性信息结构基础上的智能进化，则是作者将于明年出版的《智能原理》一书讨论的内容之一。

4.9.3 创建信息处理新逻辑

处理能力的提升是记录信息增长的重要条件。自从20世纪40年代中期诞生电子计算机以来，记录信息的处理工具以摩尔定理的速度，持续快速增长，总处理能力的增长还远远快于摩尔定理本身。但是，互联网和数字化信息感知、采集设备导致的记录信息量增长速度更快，与计算机诞生以来几十年科学计算、事务处理模式不同的需求已经产生并正在快速增长，创建信息处理新逻辑正成为信息增长的有一个需要突破的环节。计算模式变革的动因

来自四个方面，以匹配为主要形式的极短逻辑链计算、语义计算、内生计算和离散或多中心计算。

极短逻辑链计算是相对于长逻辑链计算而言。一项本质上不可并行的计算任务的逻辑链长度对计算模式影响极大，从芯片和计算机系统的体系架构到操作系统，都因此而变，x86 芯片与 ARM，视窗操作系统与安卓系统的不同，最根本的原因就在于此。当然芯片和操作系统的性能优化还要根据主要应用的处理模型。从科学计算、事务处理到互联网应用，逻辑链环持续缩短。随着信息系统集中的信息越来越多，有一些应用正在从逻辑密集走向信息密集，如自然语言处理、语音识别、搜索、信息结构显性过程中的结构匹配、音视频信息的模式识别，等等。所谓信息密集，就是通过集中大量的信息，通过匹配来达到预期目标。这种模式几乎不需要串行计算，高度并行，对传输速度要求高、存储能力要求高、计算能力要求低。今天已有的芯片、计算机系统、操作系统都没有针对这种模式进行设计和优化。

迄今为止，除了类似“沃森”的深度学习、深度分析，具有一定语义计算的性质，所有的计算和传输，其操作对象都是信息的一种特殊符号，即 bit，而不是含义，一些最后结果表示为含义的实质上是将结果的 bit 经过一个转换过程变成为 mit，而不是对 mit 的计算。真正的语义计算是对 mit 的计算。mit 的计算应该基于显性信息结构，操作的对象是结构中一个完整的语义部分，得出的结构也是直接的语义成果。这类计算将伴随信息结构显性深入而发展、成熟。

所谓内生计算是指进行该计算的计算机系统没有因该任务的执行而输入新的信息，计算结束后也没有新的结果输出系统。在现有的信息系统中，也发生一定数量的内生计算，如数据库建立索引、备份等，但这些内生计算应该属于其他计算的附属计算，不是严格意义的内生计算。严格意义的内生计算是指一类独立的任务，如显性信息结构形成过程中，不经与人交互环节的工具完善；智能系统根据确定的策略和模式，基于已占有信息的学习，计算的成果保留在系统的相应部分。

离散计算是指分散的、近乎无中心的计算。过去十多年来，集中计算成为一种潮流，云计算更将原来服务于不同系统和目标的计算集中了起来。集中计算确实节约了计算资源，提高了计算效率，但对于信息结构显性及智能

系统的进化，集中计算可能成为整个计算架构中的很小部分，主要的计算将以分散的模式进行。

与信息结构和信息增长相关的计算逻辑需要如何创建，这是下一章的主要内容。

第 5 章

信息的逻辑

信息的逻辑是理解信息增长和信息空间形成和发展的钥匙。不同状态的信息，信息的载体和外壳，除具有物理和生命特征的相应符合物理和生命规律外，均按照信息运动的规律发生和发展，信息运动规律中非物理和生命的部分，或者说，信息运动中含义的运动规律，我们称之为信息的逻辑。

5.1 信息逻辑的架构

在第 1 章中，已经讨论了信息这一类特殊的客观存在。在信息存在的不同状态、发展的不同阶段、构成的不同部分，遵循不同的发生和发展规律，既有物理的，又有生命的，还有信息的。而遵循不同规律运动的信息，又以统一的信息运动规律决定着信息增长和自成一体、独立的信息空间形成和发展。信息含义的运动规律是信息运动的规律，也就是信息逻辑。

5.1.1 信息按不同规律运动的起源

信息按不同的规律运动，这是特殊的现象。生物体既按照生命规律运动，又受物理规律的约束。信息不仅同生物体一样，按照物理和生命的规律运动，还具有自身特殊的运动规律。这是因为信息这个客观存在的特殊性。第 1 章揭示了这一特殊性，那就是载体、外壳、含义三位一体。

载体在信息构成中所起的作用就是物理承载体，承载外壳及含义，所以它的运动完全遵循物理规律。但是，从信息的解读看，遗传信息和认知信息的载体只有在生物体内才有意义，离开生物体，它就是一个研究对象或其他性质的独立存在，不再具有遗传信息或认知信息的功能。因此，遗传信息和认知信息的载体既服从物理运动规律，也遵循生命运动规律，而自在态信息和记录态信息的载体仅遵循物理运动规律。

外壳在信息的构成中十分特殊，有时候它与载体同一，有时候，它又是含义本身。外壳不能单独存在，假定独立的外壳存在，它就不是信息构成中的外壳，而是独立的物理客体。所以，外壳的运动规律在不同态信息或信息运动的不同阶段，不同观察视野，表现出多样性。根据外壳的运动特征和观察视野，图 5.1 显示了分析外壳运动的分类。

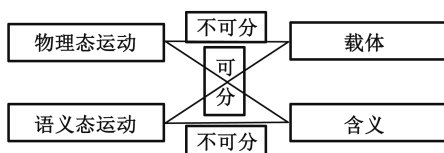


图 5.1 外壳运动分类图示

如图 5.1 所示，外壳的运动与载体、含义可分和不可分的关系形式上有八种组合。但是，如果物理态运动、语义态运动同时与载体、含义可分，没有任何意义；如果物理态运动与载体、语义态运动与含义可分，物理态运动和语义态运动与载体和语义均不可分，则没有实际意义。所以从可观察视野分析外壳的运动只有两种状态，即物理态运动与载体不可分，语义态运动与含义不可分。

外壳与载体不可分的信息运动，遵循物理规律。一个典型的例子是通信系统信息从信源到信宿的过程，计算机信息系统的信息处理也属于这一类型。无论是模拟的还是数字的，通信系统和计算机信息系统处理的对象不是信息构成中的含义而是外壳，是外壳在信息传输系统或处理系统中根据其载体，电子的物理规律运动。在传输和处理过程中对正确性的校验，也是对外壳这个客体的含义的校验，或者说是对外壳的物理特征与传输或处理目的是否一致的校验，而不是外壳中内含的含义的校验。

外壳与含义不可分的信息运动，遵循信息规律。前面分析过，即使存在外壳与载体和含义同时不可分的运动，但从解释信息运动规律的角度看，一起讨论没有实际意义，需要将两者区分开后，分别进行观察和分析。

外壳与含义不可分的运动产生于信息利用主体获取或利用信息的过程。典型的这类运动如人感知外部环境，信息传输到大脑并做出判断和响应的过程。如图 3.3 所示的光为眼球前部的色素上皮细胞感知，一路经由视杆细胞、双极细胞、无长突细胞传递到神经节细胞，再传送到如图 5.2 的中枢神经，在大脑中形成感知信息图像（或概念），并分析后大脑给出行为的指令，如回避。这个过程既是载体携带外壳的运动，也是外壳携带语义的运动。

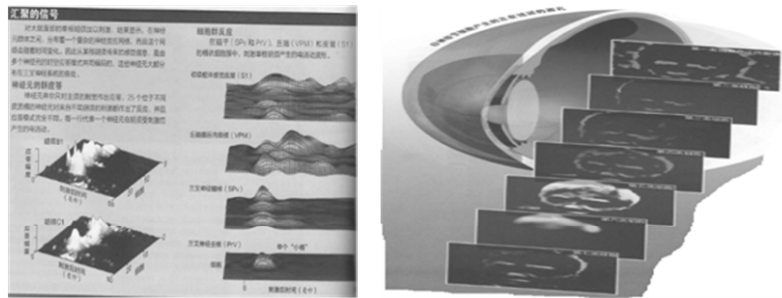


图 5.2 视觉信息的感知、传输、整合、判断运动示意

当作为遵循物理规律和生命规律的载体和符号运动过程研究时，是一个以细胞为基础，将光信号过滤分类，经过一级一级的传输，最后进入大脑的相应区域。过滤传输是电信号和化学信号，完全遵循物理规律；但整个感知和传输过程又是以生物体的生命运动规律维持的，运动过程中所有的细胞既具有信息感知或传输的功能，又具有维持生命的吸收和排泄功能。没有生命运动功能，物理运动过程是不可持续的。

当作为外壳与含义的运动时，感知和传输的同样是光电信号，但不再是根据光信号的波长或电信号的强弱打开何种细胞的阀门，而是这样的光电信号和特定细胞、神经突触的结合，代表什么样的颜色、强度和空间位置，再与大脑中相关位置已有信息的组合，形成一个关于对象的新的图像，再根据场景决定身体或语言的反应。这个过程中，所有的运动和结构，关注的都是含义，而不是载体或外壳。

记录态信息外壳与语义连接的运动方式与这两类自有态信息的情况有所不同，主要的不同点是在一般场景下，两种运动不会同时在一个运动过程中发生。记录态外壳与语义连接的运动主要发生在两种场景，主体利用记录信息和显性信息结构的生长。主体利用记录信息，不管是通过视觉、听觉还是知觉，外壳和载体都没有运动，知识含义通过相应的物理途径为主体感知，感知之后就是刚刚讨论的认知信息的运动过程了。

显性和隐性信息结构的生长都遵循信息运动的规律。这是本章主要内容。

5.1.2 信息逻辑的组成要素

信息逻辑是指信息含义的运动规律，特别是运动过程中的逻辑关系。本书第2章讨论了信息的运动，主要是载体和外壳的运动，本章将重点分析含义运动的规律和逻辑；本书第3章讨论了基于语义的信息结构，第4章讨论了信息三个构成部分不同的增长和进化过程，本章将继续讨论基于语义的信息结构，特别是显性信息结构中形成和增长的逻辑问题。

图5.3展示了与信息运动、增长和进化密切相关的四个部分逻辑构件。四个部分是感知、连接、进化和计算。其中，感知是信息运动、增长和进化的逻辑起点。感知有三类主体：生物、系统和信息结构。感知也是三态信息间转换的必要条件，基于含义的等价转换和映射是关于感知和对感知信息转换的逻辑特征。连接是关于信息结构形成和发展最为重要的功能，这里通过对显性信息结构连接的类型、逻辑关系及表述方法的分析，由内连接和外连接，实连接、虚连接和任意连接，连接实现的连接场等构件。计算逻辑部分根据显性信息结构形成和完备的计算需求和特征，从基于符号（外壳）的计算到基于含义、兼顾符号的计算，计算逻辑发生了质的变化。逻辑架构包含了以含义为基础的信息运动及显性信息结构形成和发展过程中的主要计算需求及其特殊逻辑：泛包容、微处理、短逻辑、内计算、无中心和多维计算、构件集合计算及持久扩展的显性信息结构体系架构特征，这些计算逻辑构成了新的计算模式。计算部分也包括了关于信息结构进化的一些重要逻辑问题，显性信息结构持久扩展的简约和冗余，对信息结构的存在形态及其进化逻辑进行讨论；信息结构的比较、分解和组合是显性信息结构发展过程中三种主要逻辑功能。进化部分主要由非连续的连续和具体的叠加，即从具体到具体，从已知求已知、从整体到局部的进化逻辑。非连续的连续进化则在归纳历史进程的基础上，说明信息增长的一个个具体过程充满了中断，但又如何从一个比较长的历史尺度中实现了持续的、连续的增长规律中内含的逻辑。



图 5.3 信息逻辑的构成示意图

本章以后各节，将围绕图 5.3 展开。

5.1.3 信息逻辑分析的标识方式

为本章讨论信息逻辑的需要，本节约定信息及信息结构描述的一般结构和表示方法。一般结构分三个层次，信息点、信息单元和信息结构。

信息点的构成与表述方式如表 5.1 所示。

表 5.1 信息点的构成与表述方式

构成要素一级	构成要素二级	表述方式	说明
属性描述	属性描述	$a_1, a_2, a_3, \dots a_{n1}$	信息点中所有属性描述格式
功能描述	功能总描述	$f_1, f_2, f_3, \dots f_{n2}$	信息点中整体功能描述
功能描述	感知功能描述	$fs_1, fs_2, fs_3, \dots fs_{m1}$	信息点中感知功能描述
	连接功能描述	$fc_1, fc_2, fc_3, \dots fc_{m2}$	信息点中连接功能描述
	处理功能描述	$fp_1, fp_2, fp_3, \dots fp_{m3}$	信息点中处理功能描述

其中， n_1, n_2, m_1, m_2, m_3 都是正整数。属性功能描述和功能描述的分类如有必要，均可以继续细化，属性描述从构成要素二级开始展开，功能描述继续从三级开始展开。

信息单元的构成与表述方式如表 5.2 所示。

表 5.2 信息单元的构成与表述方式

构成要素一级	构成要素二级	表述方式	说明
信息点标识	信息点标识	$p_1, p_2, p_3, \dots, p_{n3}$	对本信息单元中点赋予唯一标识符
单元共性 属性描述	单元共性 属性描述	$ua_1, ua_2, ua_3, \dots, ua_{n4}$	信息单元没有包含在信息点中的共性属性描述
单元共性 功能描述	功能总描述	$uf_1, uf_2, uf_3, \dots, uf_{n4}$	信息单元中总体功能描述
	感知功能描述	$ufs_1, ufs_2, ufs_3, \dots, ufs_{m4}$	信息单元感知功能描述
	连接功能描述	$ufc_1, ufc_2, ufc_3, \dots, ufc_{m5}$	信息单元连接功能描述
	处理功能描述	$ufp_1, ufp_2, ufp_3, \dots, ufp_{m6}$	信息单元处理功能描述

其中, $n_3, n_4, n_5, m_4, m_5, m_6$ 都是正整数。信息单元中对属性和功能的描述都是不能置于信息点中的相应描述, 整体属性和功能的描述也可同信息点的构成一样展开第二、第三级, 甚至更细的层次。

信息结构的构成与表述方式如表 5.3 所示。

表 5.3 信息结构的构成与表述方式

构成要素一级	构成要素二级	表述方式	说明
信息单元标识	信息单元标识	$u_1, u_2, u_3, \dots, u_{n6}$	对本信息结构中的信息单元赋予唯一标识符
单元共性 属性描述	单元共性 属性描述	$la_1, la_2, la_3, \dots, la_{n7}$	信息结构没有包含在信息单元中的共性属性描述
单元共性 功能描述	功能总描述	$lf_1, lf_2, lf_3, \dots, lf_{n8}$	信息结构中总体功能描述
	感知功能描述	$lfs_1, lfs_2, lfs_3, \dots, lfs_{m7}$	信息结构感知功能描述
	连接功能描述	$lfc_1, lfc_2, lfc_3, \dots, lfc_{m8}$	信息结构连接功能描述
	处理功能描述	$lfp_1, lfp_2, lfp_3, \dots, lfp_{m9}$	信息结构处理功能描述

其中, $n_6, n_7, n_8, m_7, m_8, m_9$ 都是正整数。信息结构中对属性和功能的描述都是不能置于信息单元和信息点中的相应描述, 整体属性和功能的描述也可与信息点的构成一样展开第二、第三级, 甚至更细的层次。

三个层次表述的整体以表来体现，而表述方式的每一行，包括更细层级的每一行，都是一个集合，但集合的构成和特点并不总是相同。

5.2 感知、转换和映射

本节讨论信息含义运动的规律和逻辑。首先依据第 2 章信息转换和信息移动的一般模式，分析在信息转换和移动中信息含义运动的规律，然后分别对感知、转换、移动中接收方信息结构变化进行讨论，并对 2.2 留下的满足度、通过率、转换率等逻辑问题给出解答。

5.2.1 信息含义在信息转换和移动中的一般规律

在第 2 章，信息转换定义为：一类信息形态的任意信息集合 A 经过一个转换过程变成另一类信息形态的信息集合 \hat{A} 的过程，即 $A \rightarrow \hat{A}$ 。其中 \rightarrow 表示实现这一次转换的转换过程。并依据这一定义，提出了 6 种信息态之间的转换模式即自在态信息向自有态信息转换、自在态向记录态转换、自有态向记录态转换、自有态向自在态转换、记录态向自有态转换、记录态向自在态转换。图 5.4 是图 2.1 的改编版，把载体和外壳的形态转换替代为含义的感知，并增加了显性信息结构间的感知。

在这里，“感”与“知”作为感知主体的两个阶段看，先是“感”，即感知主体与被感知客体的“接触”；后是“知”，即感知主体得到或没得到新的信息，这里的得到可以是原来没有的，也可以是与原来不同的。感知主体有 3 类，生物体、信息系统和显性信息结构；被感知客体是任何为感知主体感知而获取信息的对象。一般意义上，“感”的对象当作物理客体，“知”的对象当作含义客体，尽管两者经常是同一客观存在，这也是理解信息由载体、外壳、含义构成的一个重要途径。

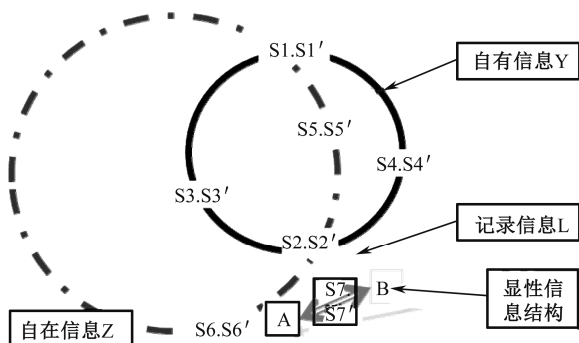


图 5.4 三态信息转换与显性信息结构间的感知边界示意图

在信息的含义这个构成要素的意义上，感知和转换是同一个过程的两个阶段，这一过程如图 5.5 所示，第一阶段是主体在一定的意愿、目的或环境的触发下，对特定的客体进行感知，感知经由特定的通道并经过目的性或功能性过滤，进入主体；第二阶段是主体对经过过滤的信息进行分析、判断、组合，进入到主体已存的隐性或显性信息结构中，如有必要，采取适当的行动，不管是否采取行动，将进入下一轮的感知或转换过程。

图 5.5 中，左边一排是各类客体，一般情景下，显性信息结构也是记录信息，但其感知和组合具有特殊性，所有单列。中间一排是三类主体，它们的两边是前后两次的过滤或处理过程。在判断或组合过程后是相应主体的信息结构增长，如果需要做出反应或行动，则同时执行，然后进入下一次感知过程。第 2 章关于信息态间一次转换是信息集合 $A \rightarrow \hat{A}$ 的过程，则对于含义的感知和转换则 A 和 \hat{A} 都是关于信息结构的一次映射过程。本节将分别对生物体和信息系统的主要感知、转换类型根据这一抽象过程作简单的分析，显性信息结构的感知及后续过程，放到本章后几节讨论。

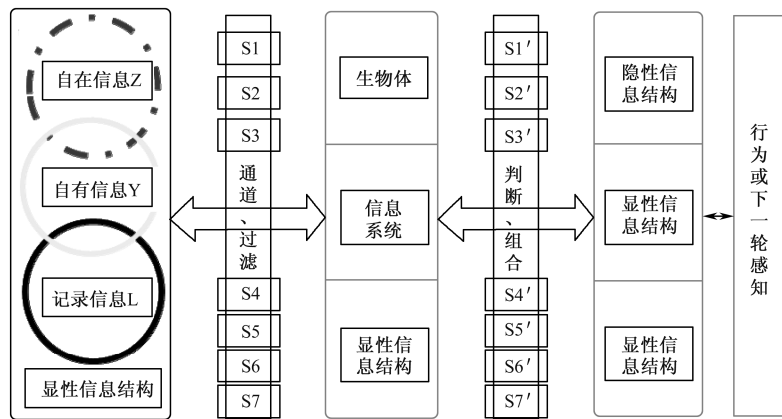


图 5.5 信息感知和转换过程的抽象

5.2.2 生物体的感知和转换

所有生物体可以感知自在态信息，部分生物体可以感知自在态和自有态信息，人类可以感知自在、自有和记录态信息，具有一定知识水准的人类可以感知图 5.5 列举的所有客体。

生物体可以感知的自在态信息取决于自身的感知能力和环境决定的可感知性。假定环境是可感知的，那么任意一次具体的感知过程，感知能力成为决定性因素。不同的生物体具有视觉、听觉、味觉、嗅觉、躯体知觉等知觉的一种或多种，灵敏度也各不相同。生物体对自有态信息的感知取决于一个可接触的群体中是否存在可相互交流的信号系统，如声音、文字或肢体语言（包括味道）。生物体对记录态信息的感知取决于在一个群体中是否存在相同的文字及文字构成的概念体系，目前，只有人具备对记录信息的感知。而显性信息结构的感知就只有具备一定知识水准的人类了。根据这样的分析，生物体感知和转换信息的过程就以具备最完整的感知能力的人为代表。

人感知各类客体，通道就是五种知觉系统。五种知觉系统并不能感知各类客体的所有的信息。对于自在态信息，依赖于知觉系统的能力，例如视觉

只能感知可见光,只能在视觉可及的范围内,其他四种知觉也是这样,这就是对自在态信息的第一次过滤:知觉信息将超出感知范围的信息统统排除在外。对于自有态信息的感知,就是人与人之间不经由记录信息的直接交流,依赖于交流环境和交流双方的默契程度,所谓默契程度就是语言、概念、交流主题的理解和认同等。而所有的不默契,或交流过程中的其他噪声,就成为自有态信息感知的过滤。对于记录态信息,感知的基础在于认知能力或知识水平,对主题是否感兴趣等主观因素,而这些主观因素也就成为对记录态信息感知的过滤。而对显性信息结构,理论上具备一定的知识水准就可以感知,但实际上,由于显性信息结构如此枯燥,除非特殊目的或研究工作需要,很多人不会化时间去感知并吸取显性信息结构中的含义,来补充自身信息结构的不足,尽管客观上这是一种很好的补充方式。

在经过第一次过滤之后,对进入神经系统的信息,还要进行第二次过滤。第二次过滤有两个原因,一是针对任务的筛选,二是针对被动感知的筛选。这里的任务是广义的使用,指所有有目的的感知,或者说所有主动感知。相对于任务的筛选,只将对任务执行有用的感知信息与任务相关的信息框架结合起来,其余的放弃;相对于被动的筛选,只把主体感知后兴趣点所在的信息组合到已有的相关记忆信息框架中;主动感知的信息在于任务结合的部分使用过后,感知的信息也同被动感知的一样,根据主体的兴趣而进入记忆。从记忆的角度看,人对已获取信息的取舍是一个复杂的认知心理—生理过程,有些即时忘记了,有些是短期记忆,有些进入长期记忆。

忽略这样的进程和取舍,从信息的角度看,是从已感知的信息集合 \hat{A} 到作为任务执行的信息集合 A' ,再到或直接到作为记忆的信息集合 A ,都是一次结构映射,可以记为:

$$\rightarrow A' \rightarrow A, \text{ 或 } \hat{A} \rightarrow A \quad (5.1)$$

假设 \hat{A} 的为一信息结构,其中包含 5 个信息单元 ($U_i, i=1\sim 5$),每个信息单元由 5 个信息点 ($p_i, i=1\sim 5$) 组成,每个信息点有 5 个描述 ($d_i, i=1\sim 5$),则 \hat{A} 可以表示为:

$$\left\{ \begin{array}{l} U1:[Ip1, Ip2, Ip3, Ip4, Ip5] \\ Ip1:[1d1, 1d2, 1d3, 1d4, 1d5] \\ Ip2:[2d1, 2d2, 2d3, 2d4, 2d5] \\ Ip3:[3d1, 3d2, 3d3, 3d4, 3d5] \\ Ip4:[4d1, 4d2, 4d3, 4d4, 4d5] \\ Ip5:[5d1, 5d2, 5d3, 5d4, 5d5] \\ U2, U3, U4, U5, 略 \end{array} \right\} \quad (5.2)$$

假设与该感知任务信息集合 A' 有 10 个信息单元, 其中 5 个与 \hat{A} 的 5 个相同, 但与 $U2, U3, U4$ 和 $U5$ 的信息点及描述也相同, 但有一个信息点 \hat{A} 中没有, 另有 3 个信息点的描述不同, 则 A' 可以表示为:

$$\left\{ \begin{array}{l} U1:[Ip1, Ip2, Ip3, Ip4, Ip5, Ip6] \\ Ip1:[\underline{1d1'}, 1d2, 1d3, 1d4, 1d5] \\ Ip2:[2d1, \underline{2d2'}, 2d3', 2d4, 2d5] \\ Ip3:[3d1, 3d2, \underline{3d3'}, 3d4', 3d5'] \\ Ip4:[4d1, 4d2, 4d3, 4d4, 4d5] \\ Ip5:[5d1, 5d2, 5d3, 5d4, 5d5] \\ Ip6:[6d1, 6d2, 6d3, 6d4, 6d5] \\ U2, U3, \dots, U10 略 \end{array} \right\} \quad (5.3)$$

其中, $Ip6$ 是不同的描述点, $Ip1, Ip2, Ip3$ 中划线的描述是与 \hat{A} 不同的。

假设与该感知相关的记忆信息集合 A 有 50 个信息单元, 其中与任务相关的与 A' 完全一致, 则 A 可以表示为:

$$\left\{ \begin{array}{l} U1:[Ip1, Ip2, Ip3, Ip4, Ip5, Ip6] \\ Ip1:[\underline{1d1'}, 1d2, 1d3, 1d4, 1d5] \\ Ip2:[2d1, \underline{2d2'}, 2d3', 2d4, 2d5] \\ Ip3:[3d1, 3d2, \underline{3d3'}, 3d4', 3d5'] \\ Ip4:[4d1, 4d2, 4d3, 4d4, 4d5] \\ Ip5:[5d1, 5d2, 5d3, 5d4, 5d5] \\ Ip6:[6d1, 6d2, 6d3, 6d4, 6d5] \\ U2, U3, \dots, U50 略 \end{array} \right\} \quad (5.4)$$

比较式 5.2, 5.3, 5.4, 差别产生在 $1d1, 2d1, 2d3, 3d3, 3d4, 3d5$ 及

1d1', 2d1', 2d3', 3d3', 3d4', 3d5'之间。如果这两组完全不同, 或不需要判断即使冗余或错误也可以合为一个新的信息结构, 那就做一次合并的操作就可以了, 得出形似式 5.5 的信息集合:

$$\left\{ \begin{array}{l} U1:[Ip1, Ip2, Ip3, Ip4, Ip5, Ip6] \\ Ip1:[1d1, \underline{1d1'}, 1d2, 1d3, 1d4, 1d5] \\ Ip2:[2d1, 2d2, \underline{2d2', 2d3, 2d3'}, 2d4, 2d5] \\ Ip3:[3d1, 3d2, 3d3, \underline{3d3', 3d4, 3d4'}, 3d5, 3d5'] \\ Ip4:[4d1, 4d2, 4d3, 4d4, 4d5] \\ Ip5:[5d1, 5d2, 5d3, 5d4, 5d5] \\ Ip6:[6d1, 6d2, 6d3, 6d4, 6d5] \\ U2, U3, \dots, U10略 \end{array} \right\} \quad (5.5)$$

但是, 对于 A' , 必须对这 10 个不同的信息描述做出判断, 否则行为就可能出错, 对于 A , 是一个基于主体认知功能和偏好, 以及外部环境的淘汰过程。所以, 对于 A' , 需要先进行一次对错的判断, 才能决定取舍, 而且这一判断对于 A 关于这一次感知的结果, 也会产生影响; 对于 A , 则是认知过程的淘汰。这一过程就是图 5.5 的第二次过滤。

生物体一次感知和信息态转换, 产生一个或两个结果。有行为目的或需要行为的感知产生两个结果, 先是为行为提供了信息, 然后是改变或产生了记忆。被动感知且不需要行为的产生一个结果, 就是改变或产生了记忆。对于后者, 是指本次感知的信息在原来的记忆中没有相关的存在, 全部增加; 而前者则是存在相关记忆, 修正、参考或增加。对于信息来说, 其实只关注与记忆相关的结果, 它改变了主体的隐性信息结构, 并可能对记录信息的改变做出了铺垫。

5.2.3 信息系统的感知或转换

信息系统的感知及转换已经成为信息增长的重要来源, 更是人类社会各种活动的依据或基础。从抽象模式看, 如图 5.5 所示, 信息系统的感知和不同态信息的转换与生物体相同, 但具体的行为存在重大差别。首先, 信息系统

的感知和转换都是明确的目的性，尽管在有的系统感知到什么可能出乎意料，但目的是清楚的，不存在被动感知。其次，信息系统的感知都是通过传感器，感知什么，感知到什么程度，都是预先设定的，所以第一次过滤是确定性的，不管环境如何变化，只感知传感器范围内的。再次，感知的信息是结构化的，进入到设定的系统功能中，这一系统的功能及其对信息结构化需求也是设定的。最后，信息的第二次过滤，即那些有用，那些保留，保留多长时间、保留的颗粒度、详尽度等也是已经设定的。分析图 5.6 的例子，可以更加形象地把握信息系统感知的特征。



图 5.6 以传感器为基础的自动化鲜花生产大棚¹

这样的鲜花生产由一个自动化信息系统控制，而控制的基础是各类传感器。主要的传感器有五类。第一类是空气相关指标传感器，如温度、湿度、气压、光照等；第二类是土壤相关指标传感器，如温度、湿度、氮磷钾的含量等；第三类是关于植物的相关指标，根系、茎叶、花果的生长状态，病虫害的情况等；第四类是与耕作机具相关的，如农机具行走控制的起点、终点、障碍，作业（喷药、施肥、采摘等）控制过程农机具与大棚设施交互感知等。信息系统的功能与传感器的类型、空间分布与数量，都是以特定鲜花生长的需要和自动化的程度而确定的，目的性十分明确。传感器不能缺，也不能多，感知信息的范围必须在设定的范围内，因为每个传感器连接到系统的一个具

¹ <http://tieba.baidu.com/p/2010861289>.

体功能模块，如大棚空气温度传感器连接到信息的大棚温度控制模块，这个模块根据事先设定的阈值进行调控，超过范围，它将不知如何判断决策。所以信息系统对所感知和转换信息第一阶段的过滤，是根据系统事先设定的功能结构、信息结构进行的，是确定性的。

系统接受到的自在态信息在传感器端或信息系统的接收端转换为记录态的数字化信息，并根据信息的数值，判定是否要采取动作。如果要有动作，则根据设定的程序执行。

系统对已经感知到的信息根据确定的规则和程序废弃或保存到数据库中。保留什么、废弃什么、保留多长时间都是预先设定的。为了通过数据分析，掌握鲜花生长和培育的规律，需要系统保留较长的时间系列数据。所以，信息系统对感知信息的第二阶段过滤是预先设定的，保留的信息类型、格式、时间跨度等是高度结构化的，均是预先设定的。

这是一个比较简单的例子，更加复杂、要求实时动态决策的信息系统对感知的精度和时间要求更强，处理过程中也会存在动态计算需求，但对本节讨论的主题，结论是相同的。

5.3 连接场及其表述

连接是理解信息和信息结构的钥匙。客观存在的属性、功能、过程及其在自在态、自有态、记录态信息中的反映，以及这些信息在转转换和运动过程中，是否与客观存在的真实保持一致，分析、判断、评价的基础在于有一个抽象、显性的信息结构作为基准，而这个基准的核心是包罗其所有构件、所有层次、所有可能与不可能的连接。本节所指连接，是显性信息结构中各个层级，各个构件的连接，有时会与隐性信息结构做一些比较。本节依次讨论连接的对象、连接的形式和连接的实现。

5.3.1 连接的对象

一个连接两个端点，就是对象。连接是相互的，两个端点互为对象。端点在信息结构中的位置和属性，信息结构在主题域、信息域和全域的位置，构成了复杂的连接对象体系。下面从内连接和外连接两个侧面讨论连接的对象。

5.3.1.1 内连接

如前面的讨论，显性信息结构有 7 个层次，从范围最小到最大分别是描述点、信息点、信息单元、信息结构、主题域、信息域、全域。这七个层次的连接，按对象在结构体系中的位置，可以分为内连接和外连接。

定义 5.1: 内连接。所谓内连接，是指连接两端的对象均在同一个结构中。这里的结构是特别定义的，即 7 个层次的每一个层级的独立顶点，但并不区分顶点是否在同一层级。

我们从信息点开始讨论内连接。一个信息点由多个描述点构成，描述的形式结构如图 5.7 所示。

物理地地标识	语义关系标识	功能属性标识	其他标识, 0-n	描述体
--------	--------	--------	-----------	-----

图 5.7 描述点结构示意图

显然描述点内没有内连接，所有的标识和描述都是最基本的，如果不是最基本的，将会分解为另一个描述体。信息点的结构如图 5.8 所示，由若干个描述体及关于信息点的共性描述构成。

从图 5.8 可以区分出 12 类内连接。其中信息点的总描述与各描述点的连接有 6 类，分别是标识间关系、语义间关系、感知间关系、连接间关系、处理间关系、描述点内容间关系。各描述点之间的连接也有 6 类，同样是标识间关系、语义间关系、感知间关系、连接间关系、处理间关系、描述点内容间关系。这十二类连接数量众多，在三大功能中，连接功能在信息结构中所占比重应该最大。

信息点与描述点和描述点与描述点之间名称相同的连接，不能归到同一类中。分析基于含义的连接和基于载体或外壳的连接是不同的。含义的连接

就是一条连线两个端点、两个对象间的关系。在数学或关系数据库中适用的一对多、多对多、多对一，或网状连接、树状连接，一串 n 个点的连接，对基于含义的信息结构来说，只有将含义作为外壳时的计算有意义，而对显性信息结构及理解含义间的连接没有意义。

U I p1: 1d1,1d2,1d3,1d4,1d5 (语义关系标识之一)				
物理地址标识	语义标识0-n1	功能标识0-1'	属性标识0-n1''	信息点描述体1-i
U I . P1.1d1	语义标识0-n2	功能标识0-2'	属性标识0-n2''	描述体
U I . P1.1d2	语义标识0-n3	功能标识0-3'	属性标识0-n3''	描述体
U I . P1.1d3	语义标识0-n4	功能标识0-4'	属性标识0-n4''	描述体
U I . P1.1d4	语义标识0-n5	功能标识0-5'	属性标识0-n5''	描述体
U I . P1.1d5	语义标识0-n6	功能标识0-6'	属性标识0-n6''	描述体

图 5.8 信息点结构示意图

含义与含义间的连接并不是两个数字、两个点的连接，每一个都具有特殊的意义。两个描述体的连接，如图 5.8 中 1d1—1d2 处理功能的连接，与 1d3—1d2 处理功能的连接可能是不同的处理，1d4—1d5 描述体的连接与 1d3—1d5 描述体的连接，可能指向描述体中不同的部件。同样，两个相同对象两次相同功能的连接，很可能导致不同的结果。因此，作为含义的连接，与计算、传输，甚至数据库中的两个对象的连接关系具有本质不同。计算和传输的对象是信息的外壳，不是含义；数据库中的数值型对象是定值，关系也是确定的关系，兼具含义和外壳的属性，当数值存在数据模型（或元数据等其他赋以含义的概念）解释时，该数值就是含义，不存在数据模型解释时，该数值就是外壳，数据库中文本型对象，处理的也是信息外壳，而不是含义。

信息单元的内连接除了单元内各信息点的内连接，还有信息点间和信息单元整体描述与各信息点的连接，还存在信息单元整体描述直接与各信息点的描述点连接。

如图 5.9 所示，一个信息单元内，所有的构件，每一个信息点及其描述体，都在信息单元内重构。以连接为中枢，把单元中所有描述和功能连接起来，形成一个适应各种可能发生变化、应用和增长需求的生长性信息单元。图中最下方部分截自图 5.8。图 5.8 是信息点的结构形态的示意，以表的方式，顺序表示点内的各个构成部件。信息单元如同生物体的细胞，是信息结构中的核心构件。信息单元的内连接以连接通道（图中的 C）为主干实现。

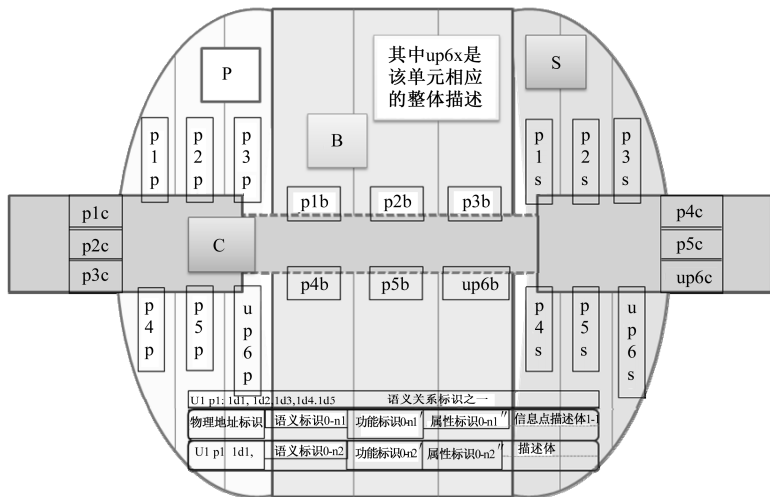


图 5.9 信息单元结构示意图

信息结构的内连接除了结构内各信息单元的内连接，还有信息单元间和信息结构整体描述单元与各信息单元的连接。同样，由于所有两个对象间的连接都是新的连接，还存在信息结构整体描述单元直接与信息单元的构成部件的直接连接。连接的实现方式见 5.3.3 节。

主题域的内连接除了域内各信息结构的内连接，还有信息结构间和主题域整体描述信息结构或单元与各信息结构的连接。同样，还存在主题域整体描述信息结构或单元直接与相关信息结构所有构成部件的直接连接。连接的实现方式见 5.3.3 节。

信息域除了域内各主题域的内连接,还有主题域间和主题域整体描述信息结构或信息单元与各主题域的连接,这里应该指出,信息域内连接相比于主题域内连接要弱,不同主题域间无论是语义、问题求解等任务,还是显性信息结构的完善,跨主题域均比主题域内的连接少。因此,信息域的内连接与前述所有范畴的内连接相比有两个显著特征,一是5.3.2节中所讨论的虚连接和无连接的数量比重大,二是在实连接中,高层次的概念、方法、功能连接比较多。连接的实现方式见5.3.3节。

全域的内连接是特殊类型,不能与其下各个层次的内连接比较。首先是全域的内连接与外连接是相同的,而其他各个层次内连接与外连接存在相当的差异。其次是除了各信息域的内连接外,信息域间连接的主要构成部分是语言和概念的理解,也就是不同语言体系的转换,也可称之为同一含义不同外壳的转换。不同语言体系的转换是一种特定的社会功能,是一种客观存在,这种客观存在映射到特定的信息结构,成为一种新的主题域。这种主题域成为全域内该类连接的主要承担者。

5.3.1.2 外连接

定义 5.2: 外连接。所谓外连接是指连接两端的对象分属不同的结构,结构的定义如定义 5.2。外连接与内连接一样都有 6 层,但内连接没有范围最小的信息点描述体,外连接没有范围最大的全域。因此,外连接涉及描述体、信息点、信息单元、信息结构、主题域和信息域。

这里应当指出的是,各层的外连接与上一层的内连接的关系。上一层的内连接中,除开下一层各构成体自身的内连接外,确实是下一层各构件的外连接,但只是这些构件的外连接的一部分,不是全部。

这里以信息结构和信息单元这两层的内连接和外连接的关系来说明,其他层次之间的关系是相同的。

如图 5.10 所示,是信息结构 A 的连接示意图。信息结构 A 有 U1~U6 共 6 个信息单元和一个整体描述单元 AU。所有信息单元均通过连接功能与内环和外环连接。图中所有与内环相接的就是信息结构 A 的内连接,而所有与外环连接的就是外连接。

从图 5.10 中我们可以清楚地看出,信息结构 A 的内连接不等同于各个信息单元的外连接。每个信息单元的外连接有两个连接点,一个通向内环,一个通向外环,通向外环的连接就不属于信息结构 A 的内连接,而属于其外连接。其他层次内外连接的关系是一样的。

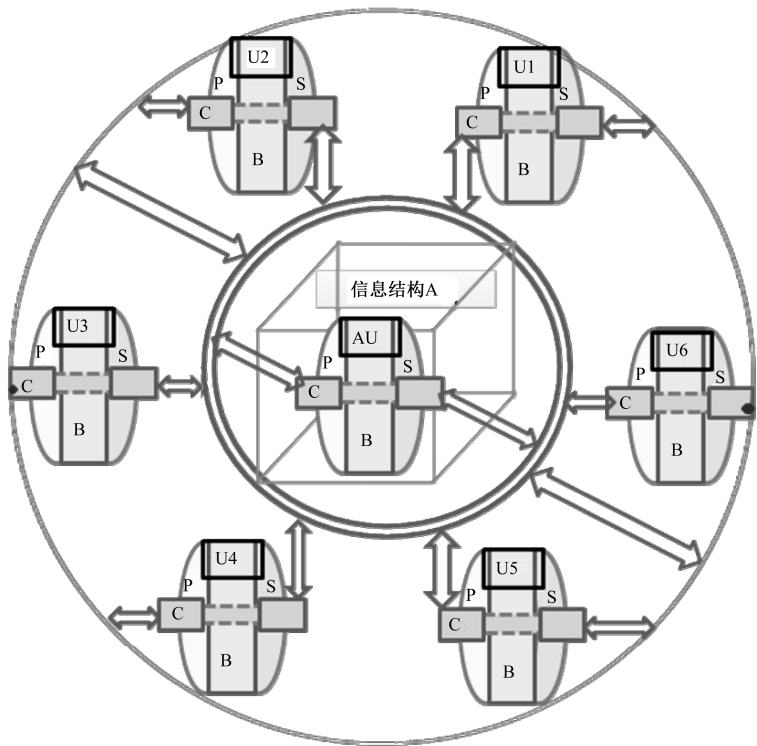


图 5.10 信息结构内外连接示意图

图 5.10 中的外环并不局限于一个主题域,而是直接与 5.1.3 节中讨论的连接场相通,也就是可以直接和不同主题域、不同信息域的各个层级的显性信息结构构件连接。理解外连接是把握以含义为基础的连接的关键。

5.3.2 连接的形式

在林林总总的信息结构中，存在复杂多样、数量众多、未来不可预知的连接，从连接的特征和实现的模式看，需要对实连接、虚连接和任意连接这三种形式做深入的分析。

5.3.2.1 实连接

定义 5.3: 实连接。实连接是指所有在 5.3.1 节中确定的内连接和外连接。这是一个确定的定义，因为所有实连接在信息结构的不同构件均已明确定义；这是一个动态的定义，因为 5.3.1 节中确定的连接随着显性信息结构的完善而增加或减少，但变化也是明确定义下的变化。

实连接本身的数量十分巨大，按照特点和实现的方式，又可以分成静态和动态两大类。静态实连接是指固定通道、固定条件的连接，在信息结构及其以下的四个层级中的大部分内连接及部分外连接，属于静态实连接。例如，信息点关于其所属的描述体的语义、物理标识，可以确定的感知、连接、处理功能与相关属性描述体的连接等。

动态连接也可以称为条件连接，连接的端点和描述均已确定，但实施连接的操作依据预先设定的条件。条件的形式根据实际连接的类型和状态而不同，常见的有阈值和事件触发。阈值是最普遍的方式，阈值在含义为基础的连接中，不一定是数值，也可以是给定的颜色、频率等。事件触发的方式更为广泛，可能是两个端点的某一个直接触发也可能是周边时间或相继事件触发。但作为实连接，阈值、条件触发的要求均已在描述中给定。

5.3.2.2 虚连接

定义 5.4: 虚连接。虚连接是指所有在语义、场景中按已有逻辑关系可能存在，但不是实连接的连接。实连接和虚连接包含了当前认知水平所有应该连接的连接。

按照定义，所有虚连接均没有在已存的信息结构中定义或描述，是可能发生还没有发生的连接；又是可能发生还没有发生的连接中没有得到连接的结果就能得到认同的连接，用通俗的语言说就是大家认为还算靠谱的连接。

虚连接的第一个特点是没有在结构中定义，定义的都是实连接。因此，虚连接形成的第一件任务是对此进行连接，什么条件下，那两个端点以什么目的连接，规定连接的通道、感知和计算的要求。

虚连接的第二个特点是符合一般的逻辑、常理。因此，虚连接大都源自借鉴、类比、启发等逻辑思维过程，正因为如此，所以即便是新连接，定义和描述同样可以借鉴、类比，比较规范，也容易实现。

虚连接的第三个特点是存在的时间一般比较短。要么第一次连接得到预期结果，进入实连接；要么没有得到预期的结果，几次尝试以后即放弃。

尽管虚连接比实连接工作量大，可靠性差，存续时间短，但虚连接是具有重要作用的连接。如果连接都是已经定义和描述清晰，意味着我们对客观存在的发生发展规律都清楚了，即使是还没有发生的自然、生命、社会事件也可以预知，这显然与我们的认知水平不符。不管是用规律解释过去、现在或未来发生的事件，还是探索发现还没有认知的规律，虚连接是客观存在，也十分重要。

5.3.2.3 任意连接

定义 5.5: 任意连接。任意连接是指超出当前已收集到记录信息中存在的各类连接，并且不按照已有逻辑推理规则提出的任意两点间连接。一般而言，也是不符合常识、常理的连接，按一般认知不知道是否合理的连接。用通俗的话来说，就是不靠谱的连接，把完全没有逻辑关系、场景支持的对象连接起来。不靠谱的连接作为三种连接形式的一种，道理何在？

在显性信息结构中，实连接对已经有明确答案的显性结构构建有意义，虚连接对不太清晰的显性结构构建有意义，任意连接对不清晰的信息结构显性有意义。前面已经讨论过，显性信息结构完备对应于其所表述的客观存在，如果在显性信息结构中已经可以明确定义和描述的连接，说明与此连接相关的客观存在的本质或规律已经清晰，需要用这个已经清晰的本质或规律求解的问题就是对已经有解的问题求解，只是任务或场景的需要，而不能为信息结构的完备提供增量。经由虚连接把问题解决了，说明是已经共识的逻辑或基于逻辑的推理能够得出结论，因此不是难题。真正的难题是超越共识的新

逻辑、新推理，而这样的新逻辑、新推理首先体现在对象间的连接上。

任意连接对认知主体的个体或群体认知能力发展具有重要的作用，对显性信息结构完备过程的自动化具有关键作用，也是下一节内计算中的一个重要任务来源。

5.3.3 连接的实现

所有的连接需要一种或多种方式来实现。实连接的实现不要再讨论，因为所有实连接，哪怕是动态的实连接，其实现方式已经在信息结构中定义。本节只讨论虚连接和任意连接的实现，同时讨论作为连接的整体，三种连接如何实现整体优化。

5.3.3.1 连接场的构成和实现

信息结构连接的广泛、复杂、机理各不相同，需要一个满足各种连接需要的工具或通道，本书称之为连接场。

定义 5.6: 连接场。所谓连接场是指在信息显性结构的各个层级，各种内外连接、实连接、虚连接、任意连接实现的通道及其相应功能。显然，这个连接场是构造的而不是天然的，不是某种物理力构成的场或某种数学规则构成的场。

连接场至少有三个构件：通道、感知、连接的确立与释放。形似图 5.11。

不同的范畴对应不同的连接场，要求是内外连接全通达。分析信息结构七个层次，连接场的范畴以信息结构为基础，以信息域为边界比较恰当。描述体、信息点和信息单元三个层次的独立性较差，全域的互通存在对外壳理解的根本障碍，主体域的边界比较容易突破。图 5.11 假定了一个有两组主题域构成的信息域，以信息结构为基础的信息域连接场，共有两组 7 个外连接环衔接在一起。

由于信息域的范围比较大，一般会跨域若干个国家，所以，通道以互联网为基础，建立适合连接场需求的专用平台，实际上，连接场就是按照一定规则，将信息结构的外环连接到互联网上，形成一个工作平台。连接场平台

的主要功能一是连接，二是感知，三是建立并维护关于连接场、连接发起、感知、确立与释放的规则。

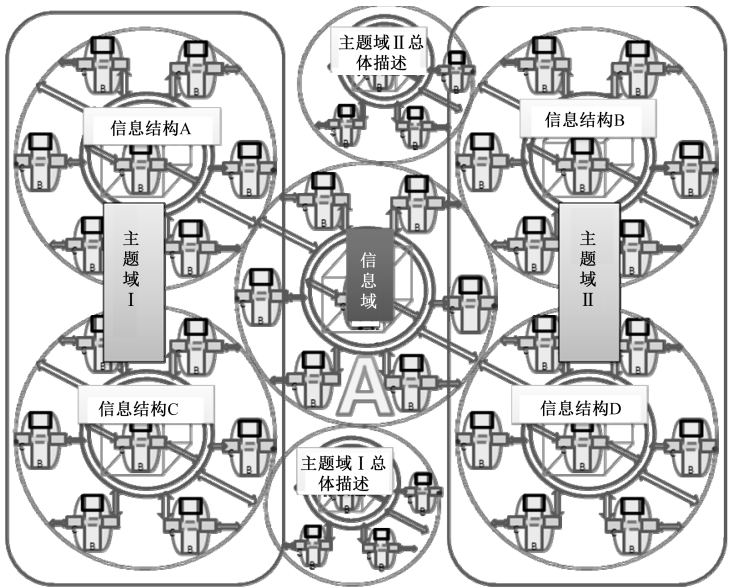


图 5.11 连接场示意图

连接主要是指信息结构以上层次的内外连接，因为一个信息结构通常为一个主体承担，信息结构以下的内连接一般在一个具体的信息系统内可以实现，而所有下层的外连接，如图 5.10 所示，都可以通过信息结构的外连接环实现。

对于实连接，感知和连接的确立与释放，以及其他关于连接的规则，都是在描述体中规定的，对来自连接场的连接请求进行分辨就是感知，分辨后符合规定的要求，就建立连接，完成连接的含义交换，就释放本次连接通道。

5.3.3.2 虚连接和任意连接的实现

相对于实连接，虚连接和任意连接的实现需要一系列规则和规则实现的途径。

在虚连接或任意连接中一个连接的建立，需要明确连接两端两个对象的

主被动关系。所谓主动的对象，就是一个连接的两个对象中发起连接请求的一方；所谓被动的对象，就是一个连接的两个对象中，接受连接请求的一方。

连接的起点是发起，要明确发起和撤回的条件。虚连接和任意连接在已有的各层次信息结构中都没有定义，规定发起的条件就是连接的基础。发起的条件有三类，一是自身结构完备需求，是指在结构完备过程中，承担主体发现利用其他信息结构的成果可能对下一步完善可以发挥作用，但又不知道具体在哪儿，提出连接的请求；二是新发现扩散，一个信息结构在完备过程中取得的新成果、新发现，包括证实的错误，认为可能对其他信息结构的完备有用，将其贡献到连接场中；三是问题征询，对信息结构完备过程中一些问题的确定存在疑虑，需求求证，于是发出请求。归纳起来，连接发起都源自信息结构完备的需求，具有明确的目的性。

按照上述关于连接发起的条件，可以设想，在连接场这个平台上，存在大量的连接请求。谁感知，如何建立连接，连接通道的释放，同样需要明确。

首先要规定一个请求的格式，连接请求格式规定的目的是方便感知和连接建立后的信息交换。一个发送到连接场的连接请求，至少满足下列条件：唯一标识码，属于三类请求中的哪一类，具体请求的特征描述。唯一标识码需要在连接场范畴内保证唯一，可以直接定位到信息结构，信息结构内的定位，通过其他构件实现。请求有三类，可以通过一个代码标识属于那一类。具体请求要给出该请求最适合层级的语义标识，能使信息场范畴内尽可能多的所有潜在接收方能理解。这三个条件，唯一标识码为了定位，建立的连接是确定的，可以根据定位交换信息，即连接的目的是可实现的。后两个是为接收方的感知服务。

对于虚连接，各个信息结构的感知功能，全天候在连接场持续搜寻与本信息结构相关的连接请求，对连接场中存在的各类请求的后两类，特别是第三类进行分辨，与本信息结构各构件及整体进展中的相关性进行比对，以决定是否建立连接，必要时可以同构造本信息结构的主体交互后决定。因此，虚连接这类感知功能属性的描述体中，包含本信息单元或信息结构在完备的当前阶段所有需要向外部求助的语义标识，通过设定的或动态设定的规则确定与连接场中哪一个连接请求建立连接。连接后进行相关信息的交换，交换

结束后，释放连接通道。

对于任意连接，一个信息结构外连接的感知功能将采用某种设定的或动态调整的随机选择方案，与连接场中的被选定的对象建立连接并进行信息交换，交换结束后释放连接通道。与虚连接不同，任意连接的接收方向发起方的提供的信息是需要确认的，要满足以下两个条件中的一个，一是对方也是任意连接的发起者，二是对方同意接受。

对于虚连接和任意连接，在交换信息后均要对得到的信息进行分析处理，做出进一步的结论。这里要指出，得到信息的一方可能是连接的发起方，也可能是连接的接收方，依赖于建立的连接是基于请求的哪一类型。

对于虚连接和任意连接，需要制定规则，保证连接场的秩序稳定，保证连接双方的利益。但对于感知的约束则应主要由相应的信息结构主体确定，使这两种逻辑不严格甚至无逻辑的连接发挥充分的作用。

5.4 计算规划和架构

以显性信息结构的完备和在既存显性信息结构的基础上进行问题求解，计算的逻辑和模式与现在讨论的计算存在本质不同，有什么不同，如何做到这类计算优化就是本节的主要内容。

5.4.1 计算规划与架构的一般性问题

基于信息结构的计算，与传统计算的不同之处就在于各自处理信息构成中的不同部分。传统计算处理信息的外壳，也就是符号，不经特殊安排，一般不是符号背后的意义。基于信息结构的计算，整体上处理的是含义，这是质的差别。但是，处理本身，即处理所利用的计算资源，又是基于处理符号的计算机系统或信息系统中进行的。要使这种计算模式在计算资源的利用和目的实现有效，必须进行规划和优化，形成基于优化的适合此类计算模式的体系架构则是讨论的重点。

传统的计算机，处理的是信息，但这个信息是负载含义的外壳。以数字计算机为例，对一串0~1的操纵，不管是计算还是搜索，表象看是含义，如计算结果的数字是“1”就不是“2”，对一个人名“张三”进行搜索，结果一定是带有张三的。但在基于显性信息结构的计算中，这个结果只能算作符合代表含义的符号（外壳），而不是含义本身。因为同样是“1”或者“2”，在不同的语境中，含义是千差万别的，不存在不同的场景中相同的“1”。对于“张三”也一样，不仅可能有很多“张三”，有男也有女，就是同一个“张三”，在不同的场景中实际角色不同，如果把同一个人姓名的“张三”作为一个信息结构中的信息单元，可能有很多个信息单元，才能真实地把“张三”在不同场景中的具体行为映射到信息结构中。此“1”非彼“1”，此“张三”非彼“张三”，这就是基于显性信息结构的计算与传统计算的本质不同。

从客观存在与显性信息结构的关系看，可以很直观地理解这种不同。客观存在不能合并，哪怕是完全相同的客观存在，如两块重量、构成元素完全相同的铁，参与人员完全相同的两次会议。作为客观存在对应的信息存在的显性信息结构也不能合并，至少两块铁的空间位置不同、两次会议的时间不同，必须用两个信息单元来描述两块铁、两次会议，不能用一个信息单元乘以2，乘以2就是外壳的处理方式。由于信息构成的特殊性，可以在保留含义独立的前提下，合并完全相同的外壳。显性信息结构存在的依据就是信息含义，而不是载体或外壳。

考虑到既有的计算资源都是基于布尔代数和冯·诺依曼体系架构的计算系统，还没有适应基于显性信息结构计算需求的计算资源，这个问题就变得更加突出。需要在计算资源本身不变的前提下对各种计算需求进行规划，形成新的计算模式和新的体系架构。

定义 5.7：计算规划。本书所称计算规划是指在一个可调用计算资源内，对基于显性信息结构计算任务的优化。优化基于任务的特点和显性信息结构计算的基本原则。

基于显性信息结构所承担的计算任务主要有两类，内计算和外部计算。所谓内计算是指计算的请求来自内部、计算的结果仅用于信息结构显性工作本身，计算使用的信息有来自外部，也有来自内部，总体看，来自内部的比

重大。我们称计算使用的信息和结果都在信息结构内部的计算为纯内计算。纯内计算在基于含义的计算中占有重要的位置。内感知、内连接、虚连接、任意连接都是产生内计算的直接原因。

所谓外部计算是指以显性信息结构为基础，完成外部提出的计算任务。这里对什么样的外部计算没有限制，问题求解、非本显性信息结构完善目的的学习，等等都是潜在的外部任务。计算任务来自外部、基于显性信息结构完成的就是外部计算任务。如此定义的计算流程如图 5.12 所示。

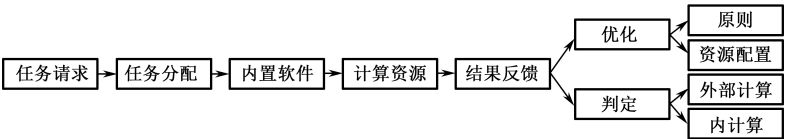


图 5.12 显性信息结构计算流程

图中任务请求来自信息结构自身完善的内计算和来自外部的计算请求。任务分配的执行主体是该显性信息结构中的计算功能系统，对于内外计算任务的安排，以内计算为主，外计算按到达时间排队。内置软件是指信息结构处理功能系统中支持各计算功能描述的软件，软件部署在计算资源，有相关功能描述体的指令调动。计算资源是该信息结构可以调用，但一般并不专属它，其他信息结构也可以调用。计算产生结果后将根据任务来源和该信息结构规定的反馈处理流程执行。后续处理一般分成两个部分。一部分是判定，外部任务将结果反馈给来源地，内计算根据判断决定是否执行信息结构内部相应构件的增加、修改、调整的目标。另一部分是根据结果，依据确定的优化准则，对相关原则和计算资源配置进行修正或不修正。

图 5.12 也可以看作显性信息结构初始计算规划或计算体系架构，在一次任务执行中不断完善。完善的功能应内置于该信息结构的处理功能系统中。

从图 5.12 中可以看出，优化有两个部分，一是软件决定计算任务后的传统意义上的计算资源优化，已有大量的研究成果和实际算法，本书不予讨论。二是如何在基于含义的计算任务通过基于符号的计算资源的过程中，两者更好地结合，这就是本节余下部分的主要内容。

5.4.2 微处理、短逻辑

不论是一个具体的客观存在，还是作为其信息存在集中体现的显性信息结构，都十分复杂，如何使该客观存在的显性信息结构真实地反映其各种属性，认知的发展是一个渐进的过程，显性信息结构的完备更是要小步走，走对一步再走一步，才能走到完备的顶点。而小步走就体现在微处理、短逻辑，把复杂问题简单化，简化到各参与主体能够理解、判断，增长就建立在可实现的基础上。

5.4.2.1 微处理

定义 5.8：微处理。微处理是一个信息结构内处理功能的部署原则，该原则明确把独立处理功能赋予尽可能小的构件（见图 5.13）。

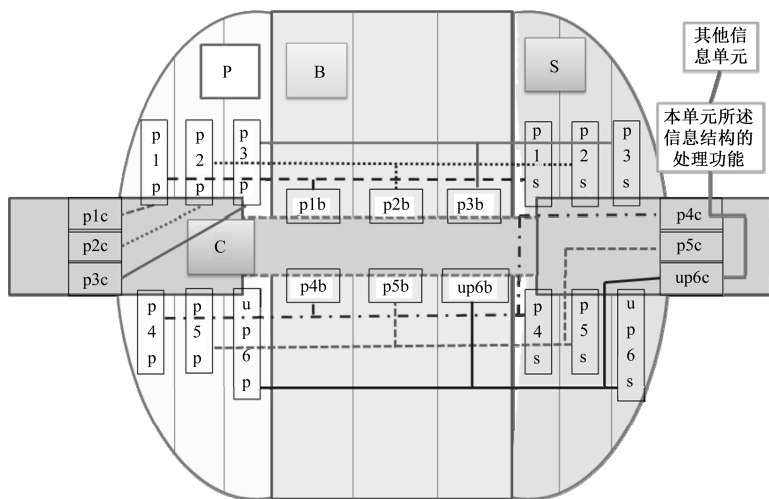


图 5.13 以信息单元为例的微处理原则

图 5.13 一幅细化后的信息单元处理功能图，处理功能分布到所有信息点的所有功能和描述，每一个感知和连接，每一个描述的发展都有自己的处理功能，通过一个个不同的功能实现感知、连接和描述体判断、增删改等信息

结构完备的所有功能。所有的微处理对应独特的软件，该软件连接计算资源，实现这个功能。由于微处理对着唯一的张三及其唯一的场景，符号处理与该符号代表的语义一致，实现了符号处理到语义处理的转换。

5.4.2.2 短逻辑

定义 5.9：短逻辑。本书所指短逻辑是指在信息结构显性的完备过程或外来计算任务的处理过程中，每一步都只推进到相应信息单元的相应构件能实现或能做出判断的位置，如果不能实现或不能做出判断，则原地停下。这个定义隐含着如果不能实现或做出前进（结构增长）的判断，信息结构完善在这个点上就会停滞，外部计算任务给对方的反馈是没有完成。停滞和没完成作为计算任务的合理结果，与长期形成的问题求解目的性产生了冲突。

其实这是符合信息结构显性完备特点和规律的有效计算策略。首先，对于外部计算任务而言，显性信息结构承担的前提是已有的结构可以得出结论，如果不能得出结论，不为外部任务承担任何义务，之所以承担，只要是证实已经取得的成果。对于信息结构完备而言，这是参与广度与逻辑深度的交换。

前面已经讨论过，信息结构完备是以当前已经取得的认知成果为依据的，超越认知水平的问题，不是信息结构完备的任务。同时，信息结构完备的计算主要是内计算，而所有的内计算都是通过基于互联网的平台，通过一定的激励机制，让所有感兴趣的网民参与。微计算、短逻辑既符合信息结构完善的规律，又适合更多的人参与。所以，只要参与的范围足够广泛，一个信息结构完备的各个部分、各个阶段总会有能够判断的主体参与进来，问题本身就能得到解决为基本假定，用参与的广度来弥补短逻辑的缺陷，就能达到预期的目标。但是，广泛的自由参与，带来的是不确定性，对于这样的不确定性，解决的办法还是把时间也算进去的参与广度。

5.4.3 泛包容

短逻辑的机理是通过广泛参与来弥补逻辑深度不足的缺陷，但广泛参与必然导致判定结论的不同，如何处理这些不同是信息结构计算的核心，原则

就是泛包容。

定义 5.10: 泛包容。本书所称泛包容是指被部分主体甚至绝大部分主体认为错误的构件,但没有形成共识之前,应该包容在显性信息结构中,而且位置不变。这个定义是信息结构增长中最重要的原则。与计算机系统或信息系统的容错不同,不是提高可靠性,而是将许多主体判定为错误的构件,依然保留在显性信息结构中,不仅保留,其位置也不因大部分人判定为错误而改变,不同的结论在结构中用并列的方式处理。

泛包容是含义计算的本质规律决定的。前面已经多次论及,含义的唯一性和正确与否的判定都是相对的,依赖于具体的场景。居于其他场景的正确性和视为合理的逻辑推理得出的结论是否正确,需要反复验证。从正确与否如何判断这个维度看,可以把信息结构增长的计算结果分成两类,一类是在当前的认知体系中有共识的,另一类是没有共识的。只要有共识,就可以假定一定可与通过广泛参与得出与共识一致的结论,从而实现与该计算相一致的信息结构增长。只要没有共识,一定会在广泛的参与下产生多个不同结论并存的局面,对于这样的状态,保留不同结论,等待共识形成,既是必然选择,也有利于加快共识形成。

泛包容原则就要容任错误、放宽约束。这与人工智能、知识工程、信息表示等相关领域力求精确无误背道而驰。其实,只要放在具体的环境中不同的原则都符合各自的目标。人工智能、知识工程、信息表示都服务于问题求解,要对系统接受的任务给出确定的回答,应该建立在拥有的知识、信息和推理过程正确无误的基础上。而显性信息结构的计算,是完成对客观存在信息映射,是客观存在世界的信息镜像,伴随着客观存在而动态发展,以智能主体的认知水平为准绳,在达成共识之前包容不同,是符合事物发展规律的,也不存在更好的选择。同时,正如前面提及,人工智能依赖的计算同信息结构显性依赖的计算一样,不是含义计算时符号或外壳的计算,要高度重视如何保证通过一系列原则、原则和模式,保证从符号计算到含义计算转换的正确性。

泛包容是与微计算、短逻辑匹配的原则。采用微计算、短逻辑作为主要计算模型,是基于泛包容的,没有泛包容,微计算、短逻辑失去基础。

5.4.4 内计算

按照 5.4.1 界定的内计算,显性信息结构完备的计算都是内计算,其中最多的又是纯内计算。在许多具有学习功能的软件中,已经存在大量内计算的例子,如字处理软件正对使用者的甄别和联想、搜索工具的个性化推送、自然语言处理中的任务处理结果的重用,等等,这些处理的共同特征是在完成一个外部任务后,根据系统内置的学习功能在后台完成一系列的内处理,在执行任务的过程中实现软件功能个性化提升。本节讨论的内计算与上述内计算有重大不同,主要是启动内计算的主要动因不是执行外部任务,而是来自内部。

5.4.4.1 内计算的主要类型和计算特点

内计算有两大类,一是直接连接显性信息结构完备的计算,包括感知、连接、描述、增长,以及相应的过程控制;二是根据结果进行后续的评价和优化,主要的对象是对整个信息结构完备所确立的 10 大原则、6 项标准和 4 类技术规范以及本章确立的与显性信息结构完备相关的原则做出评价,进行优化。

与生物体的感知不同,显性信息结构感知的实现依赖连接和计算模块。对于实连接,感知就是连接的前端,是条件符合的判断,计算就是描述体的实现。对于虚连接,感知是选择。根据感知所在构件的特点,在连接通道中选择逻辑上相关的连接请求,并实现相关内容的接受或交换,计算就是按规则进行的判断和信息接收或交换的实现。对于任意连接,感知的任务是确定一种随机选择的算法,并根据本构件当前是否具有接受任意连接的需求和能力,决定选择与否、选择什么,计算就是判断和启动随机算法。

显性信息结构的连接也是通过计算模块实现。感知是连接的前端,连接要实现是发布端的功能和通信的功能。实连接是点对点的通信,虚连接和任意连接的发布是广播式通信,而连接建立后的信息交换是点对点通信。通信功能有两种不同的场景,一是区域网,二是互联网。一个主体承担的显性信息结构完备的连接任务,一般可以在局域网实现;对不同主体承担的信息结

构间通信，一般在互联网实现。计算的任务就是实现并维护一个满足实连接、虚连接、任意连接和参与评价的志愿者工作平台。同一般的通信系统不同，这个平台的工作时间、通信峰值、参与对象是不定的，可能有规律，但规律是在实施后才能总结出来。参与对象侧应有一个计算和连接的客户端，计算的软件一般在信息结构侧，参与者在连接后获取与当前处理一致的软件客户端功能，只要任务就是给出一个需要评价的结论的判断，或提出参与者的意见。

描述的处理都在信息结构一侧，主要是描述体的规范。按照给定的标准，对已有的信息进行符合性处理，然后转入增长的处理环节。增长是确定对描述体以及，如果有必要，相对应的功能应增删改，并实现这一功能。相应的计算要求是判断和连接到结构中。判断主要依据来自内外参与者，也就是工作人员与志愿者，当发展到一定的阶段，根据判断事项的特点，可能会有计算机系统做出判断。

过程控制是对显性信息结构完备的全过程通过一个流程和程序进行控制，保证整个过程流畅、有序、规范进行。计算的任务就是编制一个实现过程控制的软件及其完善。

对于评价和优化这类计算，基础是针对已经确立的各种原则、标准、技术规范、模型、算法、流程建立经过评估的结构化软件，软件分成两个部分，服务器侧部署在计算资源中，客户端侧在参与者连接并开始工作时下载到参与者的计算资源上。

5.4.4.2 内计算的实现模式

内计算全部在机构的计算资源上实现，这里机构的计算资源可以是自身拥有的，也可以是租赁的，还可以是委托的。显性信息结构的处理功能体系不保留处理资源。

内计算实现模式实际上是指利用处理资源之外的计算功能如何实现。其实际到处理资源计算的时候，计算任务实际上已经完成，提交给计算资源的是指定软件的任务，相应的软件是信息结构处理功能的构成部分。

内计算的实现模式有两类，一类是通过确定的软件，另一类是通过信息

结构计算功能描述体内含的能调用外部计算资源微程序。

一个显性信息结构处理功能拥有的软件是一个有一定复杂程度的体系。微处理、短逻辑、内计算，使得显性信息结构的处理单元很多、很碎，但并不是要求软件一一对应。软件将相同的功能合并，将类似的功能组合，最终用几个软件来完成全部的处理需求。

这里再次出现了关键的问题，即信息结构的处理功能提交到软件，软件再与计算资源连接的时候，实际上已经将含义计算回到了符号计算，符号计算产生的具体结果重新转换为独特的含义依靠显性信息结构的功能体系，不是软件或计算资源的功能。这才是软件能合并、组合计算需求的根本原因。

很多计算十分简单，一个功能描述体将几行计算资源可执行程序包含进去，就无须经由软件这个环节了，产生了由功能体系直接提交计算的模式。显性信息结构的很多构件，尽管含义不同，但外壳（符号）相同，功能相同或相似，不同的描述体可通过标识体系调用，而不必重复这些程序的描述。

5.4.5 多中心计算

多中心计算是与微处理、短逻辑、内计算相配套的实现显性信息结构处理功能的又一种模式。

定义 5.11：多中心计算。多中心计算是指在显性信息结构完备中发生的多个计算，尽管逻辑上相互关联，但依然采用各处理中心无交互下的并行处理模式。这里的多中心处理是指显性信息结构处理功能的模式，而非委托给计算资源后的处理模式。

在一个信息结构中，如相对于人的视神经系统这一客观存在的显性信息结构，存在大量的语义和功能关联，从视网膜到大脑，逻辑关联度很深。而显性信息结构完备的基本原则又采用微处理、短逻辑、内计算，不考虑逻辑关联及其深度，需要一种处理模式架起两者间的桥梁。这种模式就是多中心的体系化处理模式。

多中心处理是一种并行处理模式，各处理中心不考虑其他处理中心与本处理中心正在进行的各处理模块之间的逻辑相关或前后系列性，直至完成计算。

不同的显性信息结构可以有不同计算中心划分方式。基本的原则是或者依据语义相关性，即信息单元间的语义关系，若干个信息单元的处理构成一个处理中心；或者是依据功能关系，相同的功能构成同一个处理中心。不管如何划分，各处理中心的整体功能中应具备一种能力，即将来自其他处理中心处理结果的通报反馈到各处理模块，并根据规定的流程，确定是否根据反馈部署新的处理。

显然，这种方式与短逻辑处理原则应对深度逻辑关联相同，用多次并行代替结果等待的串行处理。显性信息结构完备的处理需求特征，决定了采用这一系列不同于传统处理模式的必要性，而作为一种新的处理模式，通过归纳显性信息结构处理需求的特点，总结出一组原则和计算模式，来实现以含义计算和渐进式信息结构完备为核心的计算需求。

5.4.6 结构的持久扩展

本节一系列计算逻辑安排，主要原因是适应显性信息结构增长的基本规律，从计算的角度看，则是为了降低计算复杂性，提高可计算性。但是所有这些逻辑安排，从微计算、短逻辑、内计算到多中心计算，从泛包容到强制性并行计算，对显性信息结构的结构本身带来了挑战。而客观存在的无限性和认知发展的无限性，使得这一挑战更加严峻，需要对显性信息结构自身的结构持久发展做出安排。

实现结构的持久发展，首先要分辨在长期的发展过程中会碰上什么困难和矛盾。有三对主要矛盾，一是存在不可预见的增长，即在结构的什么位置、以什么速度、增长多大规模的不确定性与结构稳定性的矛盾；二是高冗余度的原则，导致存储计算的资源需求与冗余度的矛盾；三是泛包容原则，实际上的对与错并存与判断决策时构件调用策略的矛盾。下面将逐一提出解决的办法。

图 5.9 和 5.10 是简化的显性信息单元和信息结构的一个示例，我们将图 5.9 作一定的调整，就可以容纳持续、不确定增长的需求。

图 5.14 将图 5.9 中的信息点的数量、各功能构件的数量、各描述体的数

量级单元整体描述的相应部分都变成了动态的变量。因此,只要信息单元中任何构件发生变化,都可以在相应构件及其所在信息点、信息单元的描述与/或功能中得到调整。

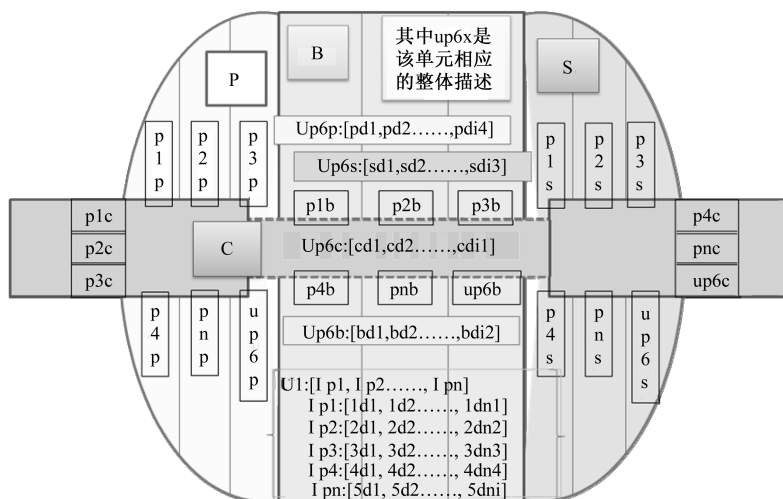


图 5.14 适应弹性增长的信息单元结构示意图

同理,图 5.10 中的各个信息单元和信息结构的整体描述作相应的调整就可以实现一个信息结构增长的弹性和稳定性的统一。

对于高冗余度与计算资源规模的矛盾，有两条解决路径。针对冗余对存储资源的压力，有很多的压缩模型和算法，这里不做介绍。从显性信息结构的角度看，保证符号与语义转换的统一性始终是高度关注的原则，主要通过实连接的地址标识实现冗余压缩后可以回归到原来的场景中，而且不采用任何导致不能完全复原的压缩算法或模型。针对压缩和指代带来的计算量增加，客观地说，相对于计算复杂性的降低，这一代价是值得的。当然也应该在信息结构显性的实践中，不断总结归纳，以发现更加有效的算法，以降低对计算资源的总需求。

针对泛包容与构件调用策略的矛盾，关键是采用什么样的策略。要从问

题解决模式下,注重得到最终结论的时间和结论的正确性转向注重当前计算的实现和计算资源的节约,而不是最终结论。在信息结构显性的过程中,最终结论有时候需要很长时间才能得出,而结构的完善是一步步实现的,即使是暂时的错误或没有进展,也不要以最终结果牺牲渐进的原则,可能损失走向正确结论的一条重要路径。

5.5 构件集合的计算和多维计算

5.4 节讨论的计算逻辑和计算对象,从信息结构完备的角度看,一般是单点生长,即使本质上是多点,也以单点增长的方式实现。但是在信息结构完备的过程中,存在一些用构件集合的方式进行计算更有效,还存在跨维度计算的需求,本节讨论这两个主题。

5.5.1 构件集合的比较

无论是在增长、验证还是审核的过程中,存在对一个个构件集合作为整体进行比较的计算需求。

定义 5.12: 构件集合。是指在一个信息结构内,以某个层次的顶点为顶点,包括其下层次的全部或部分构件组成的一个集合。包括一个层次所有的构件是构件集合的特例。参见图 5.15。

如图 5.15 所表示的是一个名为 ISA1101 的信息结构所分解出来的一个构件集合,包括了标识为 U1、U3 的两个信息单元的部分构建,各析取了三个信息点的 10 个和 8 个描述体,以及各两个单元整体描述单元的 6 个描述体。本书为叙述的简单,没有将最底层展开。在具体的信息结构中,每一个描述体应该由形如 $1d1: [x1, \dots, xi, i=1 \dots n]$ 的系列构成,其中的 xi 是具体的描述语句,是最基础的构件。

所谓构件集合的比较是指对两个或多个相关的信息构件集合进行操作,得出相互之间相同与/或相异,什么相同、什么相异的结论。有很多原因需要

进行信息结构间的比较，如增长过程，通过比较找出相异，进行后续操作，确定是否可以作为增长的新构件（构件集合）；在验证或审核过程，通过比较，可以得出是否符合标准规范、计算过程是否一致等结论。

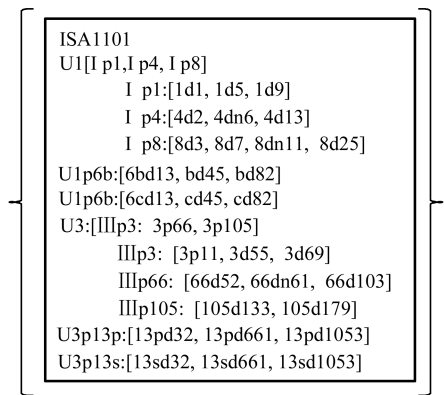


图 5.15 显性信息结构构件集合示意图

上述构件集合比较的目的存在重大的差别。作为结构增长目的的构件集比较，是以含义为基础的，而后两者，一般就是符号的一致，不是语义的一致。针对这两类比较，构件集合的比较算法是不同的。首先是析取用于比较的构建集合的方法不同。用于增长比较的构件集合，必须保证语义的可比性。保证语义可比性的根源在于信息结构所对应的客观存在的可比性。因此两个构件集合中的每一个构件，必须对应相同客体中的相同部分。这里的相同不是绝对的同一个人，而是保证比较的结论在语义背景上完全一致，可以作为信息结构完善的操作（增、删、改）。如同样分子结构，时空不同的物体，可以比较其形态或随时间、环境的变化；同一社会事件，时间不同或场景不同，可以比较关注的人或事的变化；同一作者同一主题的论文，对相同论述对象关注的角度或结论的不同；等等。

而用以验证和审核的构件比较是与标准、规范的一致性验证，标准、规范本身是归纳性的，结构的形式是其含义所在，相对于具体的显性信息结构，

实际上比较的就是包含格式在内的外壳的一致了。

构件集合的比较是一种新的计算要求，相应的在处理功能体系中要明确构件集合形成的算法和两种形式比较的算法，还要与信息增长和评价的处理要求连接。

5.5.2 基于构件集合的组合和分解

组合和分解是结构增长过程中必然存在的需求。组合和分解一定基于构件集合，单一构件组合就是 5.4 节讨论的增长，单一构件的分解没有意义。

基于构件集合的组合有两种原因，一是源于本信息结构发展过程中，发现两个独立构件集合的内容实际上反映的是同一个客观存在的相同场景，需要合并；二是源于不同信息结构在基于语义的比较中，发现反映的也是同一个客观存在的相同场景，同样需要组合在一起。基于构件集合的分解主要是因为在一个信息结构增长的过程中，发现在一个层级中存在一个构件集合，这个构件集合不属于原层级所表述的客观存在的具体场景，是同一客观存在的不同场景或表征的是不同的客观存在。

同样基于构件，组合、分解与比较的目的不同、过程不同、处理不同，但组合与分解的基础都是基于结构增长的比较及其评价结果。组合与分解的不同原因和结果，处理的过程与方法也不相同。

组合的处理相对比较简单，判断已经完成，只要将一个构件集合中不同的部分合并到另一个结构集合所在的结构中就完成了处理。

分解的处理要复杂一些。假设如图 5.15 所示的构件集合从原来的信息结构 ISA1101 中分解出，显然，它不应该是 ISA1101 中的一个新的信息单元，而是一个新的信息结构。只有分解出来的构件不能构成一个完整的信息结构，应为其中缺少了作为一个信息结构必要的构件，缺少的是信息点、信息单元、信息结构三个层次的整体描述以及部分功能。问题是，如何补齐缺失部分？是从 ISA1101 中遗传过来，还是用结构增长的方式一步步通过交互方式增长？当然，从原结构遗传是首选方案，但也需要一定的判断过程。如果经过判断，属于不同类型的客观存在的一个场景，不能遗传，则只能采用后一种方案。

5.5.3 信息结构的维度和多维计算

本章 5.3 节关于连接的讨论，特别是虚连接、任意连接和基于互联网的信息结构显性工作平台，产生了一个需要明确的新问题，信息结构的维度和跨维度计算。

定义 5.13：显性信息结构的维度。若两个显性信息结构属性描述构件间含义基本不存在可比性，或基本不存在构件集合组合或分解的可能，则称这两个信息结构分属不同的结构维度。显性信息结构维度的产生有三个主要原因，一是客观存在的多样性，很多客观存在之间语义上距离十分远；二是记录信息集合间符号体系不同，导致不能直接进行含义之间的比较；三是记录信息集合间主要概念体系不同，基本不存在属性描述构件集合的组合或分解。从显性信息结构自身的层次看，不同主题域和不同信息域均属于不同的维度。

根据定义 5.13，一般地，我们称跨主题域与/或跨信息域的计算为跨维度计算。产生跨维度计算的根源在于虚连接和任意连接，条件是存在跨主题域、信息域的连接通道。

跨维度计算带来认知瓶颈。所谓认知瓶颈是指当需要判断一个具体构件或构件集合是否能在一个信息结构相应部分增长时，由于跨维度，不管是语言的原因还是不同主题域知识水平的原因，在平台上可能出现所有参与主体不能做出正确判断的可能，也就是产生不可判定的结果。不可判定与不同主体对同一问题的不同判断不同，不同判断是有判断没共识，不可判断是无判断或有判断，但既不可否定，也不可肯定。

对于有判断无共识的场景，以前的讨论已经明确，保留不同的判断，一直到共识的形成。对于无判断或有判断无结论的场景如何决策是跨维度计算的难题，既不能简单地采取放弃的策略，这个策略可能使正确的增长没有实现；也不能简单地采取将其纳入结构的处理，因为没有依据。

根据显性信息结构增长的泛包容原则，应该采取不能证实即保留，但又不能区分策略。可以在显性信息结构中引入虚构件概念。即对无判断或有判断无结论的构件或构件集合，用一种特殊标识标记，挂接在最适合的结构位置。图 5.16 显示了一种可能的挂接方式。

图中 4 个斜体打下画线的描述点为虚构件。这样的虚构件在没有达到技术规范要求的标准做出或进入结构、或放弃的决策之前，始终保留这一状态，并与其他构件一样继续在增长流程中操作。只是与此相关的后续增长变化，均以虚结构挂接。

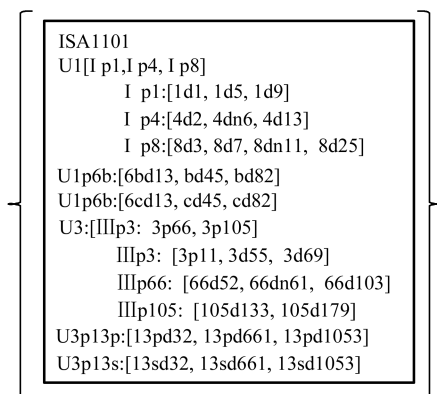


图 5.16 虚构件一种挂接方式示意图

图 5.16 为简化，只表示了描述体，其实在信息结构以下各个层次都可以采取这一方式。

当虚挂接的概念和操作方式确定后，多维计算或跨维度计算就与其他的计算相同了。

5.6 非连续的连接进化

本书第 4 章的图 4.1 揭示了信息进化的阶段，六个阶段，已经走了五个阶段，正在向第六个阶段发展，图 4.2 揭示了在自在态信息的作用下，从遗传信息到认知信息、记录信息、显性信息结构，最终形成独立的信息空间。本节要讨论是什么力量推动这一跨界、跨时空、非连续，而且多次出现重大中断的过程变成确定性的、连续的信息进化过程。

5.6.1 非连续连续进化规律

在长达几十亿年的信息增长过程中，跨越了物质、生命、信息，在各个环节曾经出现过很多次不同的中断，但信息却是连续进化，并出现了加速的迹象，我们称这一过程为信息非连续连续进化规律。图 5.17 诠释了这一规律的奥秘。

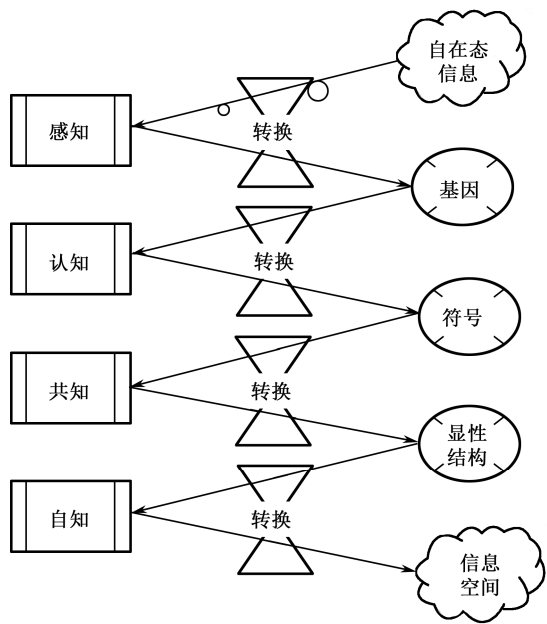


图 5.17 信息进化之路

如图 5.17 所示，经历四个阶段，八次转换，从只有存在，自身含义不能发挥作用的自在态信息进化到独立、自我发展的信息空间。

在这个进程中，每一个大的阶段遵循不同的规律，相互之间不从信息进化的角度看，都是独立的、中断的；在各个阶段之中，存在大量的中断现象，但信息演进的过程没有中断，下面逐个阶段作简要综合分析，详细的内容在后续几个小节展开。

第一个阶段有两次转换，形成两个有联系又相对独立的过程，一次是生命感知外界信息，形成遗传基因或引发遗传基因突变；另一次是对产生基因突变的生物体进行选择，向提升生物体认知能力的有益进化倾斜，一直到人类语言能力的形成。与此后的信息进化过程相比，这个阶段有两个鲜明的特征。一是时间漫长，足足走了 30 多亿年；二是完全可以由生命进化的规律来解释，似乎与信息进化无关。进化过程可以用生命演进的规律，但对规律的解释需要从信息进化的角度重新理解。

从信息进化的历程看，这是艰难的第一步。这一步要完成信息进化的两个基本功能：感知能力和认知功能，成为信息感知力，第一阶段将形成了自有态信息。为什么自然选择对这两个能力给予有益性倾斜，是生物体生存的需要，是生物体生命维持和遗传进化目的。从信息进化的角度看，似乎自然选择在强大体能和完善的认知功能之间更加倾向于后者。没有造成 6500 万年前恐龙灭绝的环境变迁，强大体能的生物是世界的主宰，有利于认知功能发展的遗传基因不会得到有益性选择。而智人将尼安德特人和其他人种全部灭绝，则应看作达尔文进化论的最佳注解之一，具有当代物种最先进的认知功能的智人，将当时适应恶劣环境、体质最强大的尼安德特人击败。尤瓦尔·赫拉利认为主要原因是智人拥有比别的生物或同类更强的语言能力²。

语言能力实际上已经是信息进化第二阶段的一个成果，认知功能已经形成了综合的认知能力，并能将认知结果用一群主体能辨识、理解的符号表述出来，成为信息表达力。这个过程超出了生命进化范畴，成为智能进化的一个环节。第二阶段两次转换形成了记录态信息。第一次是将由遗传获得的感知功能转换为人的感知能力，五感的信息在大脑汇聚，并根据外部环境做出反应。第二次是将大脑中积累的认知以语言和文字的方式表达出来，为形成信息进化由个体发展到群体创造了条件。

第三个阶段是语言和文字成为共同理解的信息，从自有信息转换为共有信息，不同主体间形成了对符号表述含义的解释能力，成为解释力。解释力

2 人类简史，从动物到上帝，[以色列]尤瓦尔·赫拉利著，林俊宏译，中信出版社，2014 年，第一部分 认知革命。

的发展也经过了两次重要的转换过程，一是主体使用相同的符号体系，且人的认知能力还能应对记录信息数量增长；二是使用不同符号体系的主体使用记录信息、记录信息数量超越人在各种具体约束下的利用能力。第一次转换时共同的符号体系的发展和完善，第二次转换是记录信息集合的结构化和基于结构化的记录信息的利用加速。

第四个阶段是以记录信息的显性结构为基础，通过更加完善的解释力，非生物主体具备理解记录信息的能力，以及自我感知获取信息、记录处理信息能力，以可以自我解释的、广泛领域的显性信息结构为基础，同时具备强大的感知力、表达力、解释力，形成独立的重现客观存在的重现力，将自在态信息、自有态信息、记录态信息包容在一起的独立信息空间。

5.6.2 感知力

如图 5.18 所示，有两种目的不同的感知力。

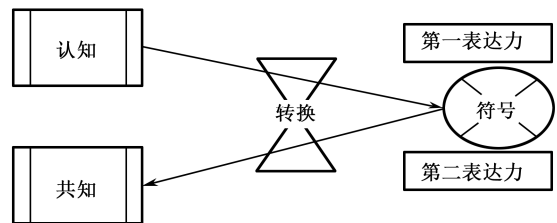


图 5.18 两种感知力

第一感知力是指生物体感知自在态信息或适应自在态信息所代表的外部环境的作用下，遗传基因在漫长的历史阶段持续突变和选择，最终使智人具备五感功能。没有遗传基因沿着感知功能方向的持续进化，形成满足认知需求的感知功能，信息的进化就止于植物或低等动物的感知能力，而不能完成最终的循环。

第二感知力是指由遗传基因形成的先天感知功能在生物体后生命过程

中成为生存的一种基本功能，并促进这种功能持续进步的实际感知能力。第二感知力对信息的进化有双重贡献，首先是感知变成基因突变朝着有利于感知功能持续完善方向进化，其次是在一生的时间内积累感知信息，朝着突破认知的第二道门槛，表达的形成发力。

感知是生命过程，遵循生命运动规律，表达则是社会过程，遵循社会发展规律，但第二感知力又与表达力形成了直接的联系。所以，认知的发展区分为感知力和表达力，感知力区分成两个阶段对于我们理解认知过程，从而为实现非生物认知主体的完整功能、走向独立、自我发展的信息空间摆脱认识论和方法论的束缚。

5.6.3 表达力

表达力是指一个（一类）认知主体将其认知结果用其他主体可以理解的外壳或符号表达出来。如图 5.19 所示，有两种形成过程截然不同的表达力。

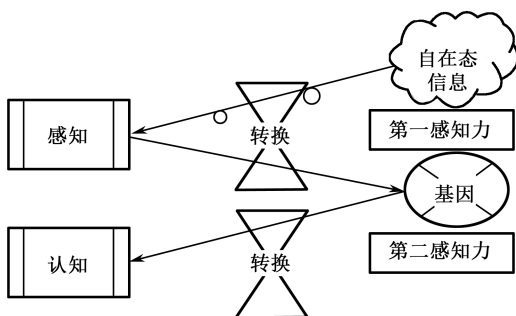


图 5.19 两种表达力

图中，第一表达力是指一个认知主体对其认识到的信息用其他主体可理解的符号表达出来的能力。第二表达力是指符号（外壳）的表达力。

第一表达力有两个核心要素，认知抽象能力和符号利用能力。这两者各

不相同但又有一定的关联。认知抽象能力是神经系统功能与社会既存的符号体系、概念体系和个体、整体认知水平共同作用的结果。符号和概念的利用能力既与社会既存符号体系、概念体系相关，更与认知主体的受教育水平和社会角色相关，不同的受教育程度和不同的社会角色等因素导致不同的符号和概念利用能力。

第二表达力是社会文明演进的结果，遵循社会发展规律。对于概念体系的演进和客观化的认知成果，支撑记录态信息在社会环境下正向生长的语言和文字发展，有不少研究成果，如周有光的《世界文字发展史》³和奥斯特勒的《语言帝国，世界语言史》⁴。尽管在上万年的历史中，存在大量的中断，但朝着丰富、互相融通的方向发展是确定的。不同信息域主体间的交流深化，必然导致植根于不同信息域的符号体系和概念体系朝着相同表达力的方向发展。一个符号及概念体系，如果不能吸纳其他信息域的表达力，不能包容不同文明及符号和概念体系的新进展，表达力不能满足相应主体集合需求，就可能进入被淘汰的行列。

从感知力发展到表达力，本质上是从生命过程进入社会过程，其进化和发展遵循不同的规律。生物体感知能力是个体的遗传基因决定的，是经由其所在的生存环境而发挥的。尽管受生存环境的制约，但根基在生命运动本身。第一表达力源自生物体，不管是该生物体是讲话还是写作，都是生命运动的一个过程。但这个生命过程更多地受社会环境的约束。一个人所处的语言和文字环境，其符号体系和概念体系的特征或成熟度，以及该主体对语言文字的掌握和运用能力，都带有深刻的社会烙印。

符号体系和概念体系的进化史一个社会进化过程，每一个具体的符号或概念，进入和扬弃，是社会的选择。对符号体系和概念体系的社会选择具有与生物进化相同的规律。1976年，英国牛津大学的道金斯在其《自私的基因》一书中，创造了“迷因”（Meme）这个词，并将这一规律归于“复制子”这

3 语言帝国 世界语言史，[英]尼古拉斯·奥斯特勒著，章璐、梵非、蒋哲杰、王草倩译，维舟校，上海人民出版社，2011年，第二版。

4 世界文字发展史，周有光著，上海教育出版社，1997年。

个社会隐藏功能⁵。“复制子”这个概念从解释延续的角度看很好用，但缺乏生命科学及社会科学的实证，不能对中断做出很好的解释。

从信息进化的角度看，表达力的形成、积聚和发展是基于人类对认知和表达的内在动力，基于社会选择规则促进着符号体系、概念体系和文明的演进。尽管不同的文明进化在局部出现中断，在选择上存在偏好，在某些领域、某些阶段对人类社会文明进步存在一定的抑制作用，这在一方面说明“复制子”在解释和实证两个方面缺乏普遍性，但又在另一方面证明了 5.6.6 节中讨论的信息进化规律。

5.6.4 解释力

解释力是指信息由符号体系这一信息的外壳转变为主体能理解的含义的能力。如图 5.20 所示，解释力也分成两个部分。

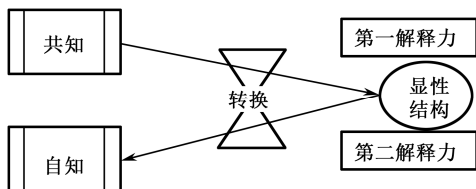


图 5.20 两种解释力

第一种解释力是将一定范围内共知的记录信息转化为显性信息结构的能力，这个过程本质上是通过结构将符号体系表示的信息含义在含义的基础上重构，利用这个结构固化共知，提升共知。例如，一本书的作者能够通晓该书所要表达的所有含义，但很多含义的各种关系、场景在写作过程中被作者省略了，这种省略对具有与作者类似知识或认知背景的主体没有关系，但对其他读者就带来了一些缺失，不利于理解，至少不利于用最少的时间理解，

5 The Selfish Gene, Third ed., [英]Richard Dawkins, Oxford University Press, 2006 年.

需要通过一种方式将这些缺失补起来。有人讲解是一种办法，但绝大部分阅读场合，不存在这样的条件，于是就有专门注解的书。如果将一个主题的作品积聚在一起，发现单独为每一本书作注解不仅存在大量的重复，而且也不一定能够达到足够的详细程度，因此，显性信息结构就承担了这个角色，成为更好地达到共知的基础。

第二种解释力是使已经形成的显性信息结构能为另一类主体能理解和利用。这里的另一类主体，一般意义上是非生物智能体，特殊意义上还有完全没有共同基础的信息域的主体，如外星人。第二种解释力就是让这一类主体在含义的基础上理解信息、利用信息，使这些主体成为已有显性信息结构的自知者。第二种解释力的核心作用是使非生物智能体跨越认知的门槛。随着传感器、物联网的发展，非生物智能体的感知功能正在不断进化，而认知功能则需要另外的途径，第二种解释力为此走出了第一步，第二步是重现力。

5.6.5 重现力

重现力是信息结构的基本功能。在隐性信息结构中，遗传基因的转录、复制，不是一个个基因简单的复制，二是一个个具有特定功能的片段按照结构的转录、复制和启动功能，这些都是隐性信息结构的重现力。在显性信息结构中，目的在于对该信息结构所反映的客观存在的重现，是根据客观存在及其动态性和功能性进行整体的表述，而不是割裂开的属性描述和功能描述。

如图 5.21 所示，重现力是从自知走向形成独立信息空间的关键环节，要通过重现力，融合信息进化的感知力、表达力和解释力，打通自在态、自有态、记录态信息的边界，形成以自知主体为核心，以显性信息结构为骨架的独立、自我完善的信息空间。

互联网、传感器、移动智能终端拓展着人类及其创造的信息系统获取信息的能力，短短几年时间，记录信息的增长超越了过去认知，大数据概念随之产生，很多人惊呼人类已经不能掌控如此快速增长的记录信息。戴维·温伯格惊呼“现如今，我们将信息超载视为一种文化环境。而令我们深夜难眠的，并非是担忧如此众多的信息会令我们精神崩溃，而是担心我们无法得到

自己需要的足够多的信息”⁶。

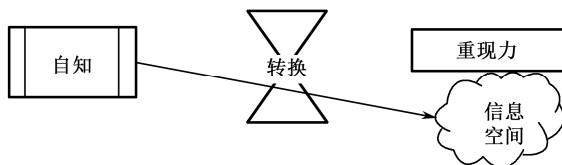


图 5.21 重现力的位置

重现力的基础是显性信息结构中的重现度。重现度是显性信息结构的一个评价标准，要求能重现其所反映的客观存在的相应场景。重现度的要求，是显性信息结构与知识表示差别的一个重要体现，即显性信息结构完备与否的参照系是客观存在，而客观存在包括自然界状态和事件、社会状态和事件、人的行为和认知，当然包括认知过程的一个子集——推理过程。

重现力超越了重现度，不仅要求能真实地重现所反映的客观存在，还要成为构建独立信息空间的主要能力。如何在重现力的平台融合感知力、表达力和解释力，有三个必要条件，一是能够实现融合的主体，二是显性信息结构覆盖的范围足够大，完备度足够高，三是能力足够强，能对一定范畴内的自在态、自有态、记录态信息自由操纵。进一步的讨论在第6章展开。

5.6.6 中断与连续

图 5.17 所勾画的信息进化之路，历经 30 多亿年，对生命演进过程和社会发展过程高度依赖，是什么力量推动着这条不断发生惊险中断的历程走到最后？

相对于认知革命的生命进化过程似乎已经走到了尽头。尽管在几十亿年中出现了多次部分物种进化的中断，但依然进化到了智人的认知水平，或者

6 知识的边界，英文书名：Too Big to Know，[美]戴维·温伯格著，胡泳、高美译，山西人民出版社，2014。

说进化到了自然演进的顶端。实际上当智人产生之后，其他物种依然在进化，但不再可能产生在认知能力上与现代人并肩的物种。赫拉利在《人类简史，从动物到上帝》一书中，多次指出智人灭绝其他人种和大型动物物种的事实⁷。如果没有智人的出现，其他类型的猿类，甚至猫、狗等都有可能进化到认知革命这一步，只是时间可能推移若干万年，若干百万年，若干千万年，但对于生命演进 35 亿年，地球还有数十亿年的生存期，百万年、千万年都不是问题。更为重要的是，从认知发展过程看，生命自然进化已经没有动力，即使是超越人类认知能力的物种出现，也是人类设计出来的⁸。因此，从信息进化与生命自然演进相关的感知力和表达力可以画上句号，不再会有重大进展。

从社会发展的角度讨论符号体系和概念体系的进化，也已经发展到不可逆的阶段。所谓不可逆，就是说不必担心中断导致文明的消亡，不必担心语言文字会倒退到不能恰当地表述人类认知的成果。一部世界语言文字的历史，一部世界文明史，都是中断和持续发展的历史。但是，所有的中断，不管是南美的玛雅文明还是亚洲的高棉王朝、西夏文化，还是欧洲希腊九里特岛上的米诺斯文明、非洲的阿克苏姆王朝，对局部语言文字的演进带来影响，但不是全局。相对于认知功能遗传基因的漫长而艰难的历程，十万年的语言史、不到一万年的文字史，进化到如此丰富多样，说明第二步表达力的发展远比感知力和第一步表达力的发展要快。

由生命自然演进和社会演进规律决定的信息进化的感知力和表达力，以及由生命和社会演进决定的解释力，不再会有重大进展，已经发展到不可逆的阶段，是否说明不会中断，我们简单分析一下信息进化感知力和表达力中断的充分必要条件。

如果具备这两个条件，信息进化的感知力、表达力和第一阶段解释力仍然会中断：一是已经演进到高级认知功能的生物体全部消亡且其他物种向认知功能进化的路径不再存在；二是所有记录全部毁灭，或虽然没有毁灭，但

7 人类简史，从动物到上帝，[以色列]尤瓦尔·赫拉利著，林俊宏译，中信出版社，2014 年，第 19 页。

8 人类简史，从动物到上帝，[以色列]尤瓦尔·赫拉利著，林俊宏译，中信出版社，2014 年，第二十章。

不可能存在能理解这些记录的主体。在一般情况下,这种条件具备,就是不再适合高等动物的生存,而且不是地球自然发展的结果。因此,可以得出这样的结论,如果不是地球因为特殊原因毁灭或不适合高等动物的生存,信息进化的感知力、表达力和第一阶段解释力不会中断。

信息进化第二阶段解释力和重现力的基本动力依然产生于人和人类社会进化的动力,那就是对认知能力和把握客观世界的愿望。如果说认知能力进化也是自然进化的偏好,那么也源于自然进化的动力。但是,信息技术的进步和社会的发展,人类要在持续快速增长的信息中得到自己需要的信息,信息越多,越担心信息不足,形成恶性循环,这就是显性信息结构之所以存在的主要理由。要在这样的信息环境中继续保持认知的领先,已经超越了作为生物体的人类的认知能力,必须借助工具,必须通过新的渠道,而迄今为止能够看到的工具和渠道就是基于显性信息结构的解释力和重现力。由此,保证信息进化全程不中断的动力得到了来源。

5.6.7 进化的时空逻辑

信息进化的不同阶段,所经历的时间完全不同,最长的时间发生在经由遗传基因进化形成的感知能力和认知能力,也就是生命进化过程中以大脑为核心的中枢神经系统的形成。在文字形成以后的几千年内,人类的认知能力快速提升,其背后的是适应着不同的进化逻辑。

在认知功能形成的生命过程中,进化是个体的,是沿着一个个物种独自前行,产生着一次又一次的变异,等待自然的选择,等待自然选择改变种群,最终具有认知功能优势的变异者胜出,认知功能进化到新阶段。串行、等待(甚至反复)、代际周期长、进展阶段多等因素合在一起,造成了不可思议的几十亿年如一日的进化,直到作为认知功能巅峰标志的智人统治地球。

而从语言到文字,逐步从串行向并行转变,个体向群体转变,少量交互向大规模实时交互转变。语言是生物体第一次有了将认知结果告知别人的载体。然而早期的语言缺乏表达认知结果、交流对象理解的有效概念体系。同时,没有概念体系和已有认知成果的学习过程,一个主体在一生很短的有效

时间内积聚的认知成果十分有限，加上交互的对象少、起点低，有没有记录实现跨越生命约束的传递，语言形成后的很长时间内，以接近于串行的方式缓慢进化。当语言形成相对稳定的概念体系，表达一定数量的客体，口口相传，将上一代、上几代积累的经验和认识传给下一代的模式制约了认知能力的发展，将认知结果变成不受时间和空间约束的记录的需求不断上升。部落规模扩大、工具的发展和同一地点居住时间变长，文字的产生在各个方面具备条件，世界几个地方在相互没有交流的情况下，在距今 5000 年左右先后产生文字，说明了这一社会进化过程的规律性。分析这几千年的历史，可以看到突破时间和空间约束的文字在认知发展中的重要作用，积累突破了个体的约束、发展突破了串行的约束，而且积累的数量越多、参与的人数和交互的平台越方便，发展的速度越快。

分析这样的认知演进过程，可以看到一个以认知事件数量、积累信息的质量、交互平台的功能为基础的演进规律，即演进速度是相应认知事件、积累信息、平台功能的函数。设演进速度为 M ， \hat{A} 、 \hat{E} 、 \hat{N} 为常数，设认知事件为 X ，积累信息为 Y ，平台功能为 Z ，有如下的关系

$$M = \hat{A}/X + \hat{E}/Y + \hat{N}/Z \quad (5.6)$$

M 是一个时间单位，如年或月、日， M 越小，速度越快。 \hat{A} 是个大整数， \hat{E} 、 \hat{N} 是与 Y 、 Z 同量纲的大数。相应认知事件是指在同一个信息域的认知主体数和一个时间周期平均认知事件数的乘积，时间单位与 M 同。平均认知事件是一个综合测度，包含了学习时间、认知整体水平等因素。

通过认知事件、积累信息、平台功能，将以时间为代价的串行进化和以空间和事件数量为基础的进化有了可比性，也为互联网普及之后信息进化速度加快提供了逻辑依据。几十亿网民在共同的平台上，只要有适当的工具和激励机制，显性信息结构演进就可能跟上信息的增长，保持认知的进步；如果自知主体全面替代人在显性信息结构完备方面的工作，则有可能超越信息增长步伐，使信息空间成为推动地球文明演进的三驾马车之一。串行进化与并行进化的逻辑转换不仅可以解释知识、指数增长的现象，还可以引申到人口数、经济发展等其他领域发展速度的变化。

5.7 信息逻辑的进一步讨论

本章从介绍信息逻辑的一般架构开始，渐次展开讨论感知、连接、计算三个功能的逻辑特征，然后讨论了信息进化过程的逻辑及其特点，我们看到了信息逻辑与其他领域逻辑的重大不同。信息逻辑与其他领域逻辑的不同基于信息的本质特征，还需要从整体对信息逻辑进行讨论。

5.7.1 连接和计算逻辑的一般问题

本章 5.3 节、5.4 节、5.5 节从不同角度揭示了信息结构显性过程的连接和计算逻辑，指出了一系列与传统连接和计算不同的特征。已经知道，之所以存在这些特征，源自信息构成和信息结构显性进程的特殊性。信息连接和计算要兼顾并区分外壳（符号）与含义的不同，显性信息结构自我演进要求的内计算、虚连接、任意连接等连接和计算要求，导致了不同的连接和计算模式。

除了判定规则的变化之外，连接和计算都是计算。计算逻辑决定于基本的控制逻辑和计算机体系架构。今天的计算体系，基于布尔代数和冯·诺依曼体系架构。这一体系对于信息结构显性所涉及的符号计算，完全可以满足需求；对于涉及的含义计算，本质上不适用。

矛盾集中在两个方面，一是自生长的计算目的，二是符号与符号所代表的含义是不同的。连接功能中的虚连接和任意连接，计算中的微处理、短逻辑、多中心计算、多维计算、内计算，都源于自生长的计算目的。在本章所有这些连接或计算模式，都变通为适应现有计算资源。问题在于，是不是存在更有效的数学逻辑和计算机体系架构？而泛包容、构件集合计算则源于符号与符号代表的含义具有不同的值、不同的量或量纲。同一个符号或相同的符号集合，在没有确定含义是否相同之前，不能进行所有基于符号相同的计

算，或者说本质上就没有含义基础上相同的符号或符号集合。同样对泛包容、构件集合计算也变通为适合已有计算资源的模式，产生的问题也相同。

差别客观存在，但是不是存在更有效的数学逻辑和计算机体系架构？到目前，只能说，应该存在，但究竟是什么，这是需要研究的课题。

5.7.2 显性信息结构自我进化的逻辑问题

从连接和计算逻辑的角度看，任意连接和泛包容是为显性信息结构自我进化设计的。但自我进化的逻辑问题还在这些具体的模式之前。

显性信息结构的参照系是客观存在，而不是记录信息或大脑中的认知信息。显性信息结构一经形成就是记录信息，但此记录信息不是彼记录信息，它已经包含了客观存在发生发展的物理或社会过程，或者决定该过程背后的物理或社会规律；它已经包含了产生大脑认知的过程或决定该过程的规律。逻辑上重现一个客观存在是相对应的信息显性结构完备的基本要求。这是说显性信息结构在起点上就决定了具备自我进化的能力。

应该说，显性信息结构的体系架构和原则、标准、规范、功能实现算法都是围绕自我进化设计的。除了原则、标准、规范的动态可扩展要求外，具备完整属性描述和处理功能的信息单元以上各层构件，微处理和多中心处理，都是从微观结构上构造了自我生长因子。

在显性信息结构的体系中，信息单元是一个核心构件，它对应客观存在一个相对独立的部分，具备显性信息结构所有的功能，感知能力、连接能力和处理能力，能够调用需要的外部计算资源。信息单元就是一个在一定的原则、标准、规范和连接、计算环境约束下的自我生长单元。

显性信息结构自我生长逻辑基于其基本构件的可生长性，并在各个重要的组成部分中配合实现，泛包容、虚连接和任意连接、微处理和多中心、多维处理都是具体的适合其自我生长需要的操作模式，是整个生长逻辑中的构件。理论上可以从最简单的显性信息结构开始逐步到达不可逆转的自我生长转折点，但如果没有前期加速的模式，到达这个转折点的时间会很长，如同生命认知功能进化产生文字前后的进化速度之差。显性信息结构发展的客

观环境比语言文字早期演进要好得多。物理环境基本具备，感知、互联网、移动终端、各类信息系统都可以成为其发展的基础设施；能够成为交互对象的主体数量众多，具备参与的各项条件。只要有机构或个体作为具体的信息结构或主题域的显性信息结构承担者，制定原则，构建基于互联网的工作平台。只要开始起步，信息增长的时间和空间、串行和并行的规律就会发挥作用，参与的人数越多，发展越快。

5.7.3 信息逻辑的基本特征

信息逻辑的特征基于信息三态的转换和三个构成部分，而在一系列特有的逻辑过程中有存在一定的共同之处。

在三态转换中，不仅是转换的目的是含义，不是载体或外壳，而且，都是具体的一对一的转换。向日葵感知阳光而转向，每一棵植株感知的信息相同，但不能合并，不能替代；遗传信息控制细胞生长，在相同的时间、相同的细胞接受相同的信息，但也不能合并，不能替代；一个班的学生听老师讲课，听的是相同的内容，得到的信息可能不一致，不管是一致还是不一致，同样不能合并、不能替代；一只脚不能两次踏进同一条河这句谚语，更说明了客观存在是随时空的变化而变化的，在信息含义的层次，没有汇总，只有具体的存在。

由于信息由三个部分构成，所以信息运动遵循的规律依据具体运动的不同而不同。载体的运动遵循物质运动规律，当外壳与载体合一时，外壳的运动也遵循载体的规律，含义的运动必须依赖载体或外壳时，遵循相应规律，而在所有以含义作为运动目的时，必然脱离原来载体和外壳而与新的载体或外壳结合运动。视杆细胞感知一支白色鸭子的羽毛后又感知一支白色的鹅的羽毛，在视杆细胞转化为电信号并传送到大脑相应部位这个过程中，两个过程的信号可能一模一样，但到大脑合成来自前后和其他信号之后，这相同的信号就会区分出鹅毛还是鸭毛，在不同的神经元处理和存储。这个例子说明，作为含义时，没有汇总，只有具体的存在。即使是同一只鸭子的同一根毛，在不同的场景有不同的描述。

上述讨论暗示在含义基础上分析信息的逻辑，其最主要的特征是从具体到具体，以具体的总和来对应在或不在总和中的具体。我们已经习惯于在把握一个具体的基础上进行抽象，总结规律，通过数学的货逻辑的方法求解同类的问题，甚至触类旁通，求解类似的问题。而信息逻辑恰恰相反，先构整体，以整体应对局部。是积累已知，结构化已知，从已知求已知。尽管有时候也会在整个过程的一些环节，使用一定的逻辑推理，但基本逻辑框架是从已知求已知。

迄今为止的数学，是关于形、数、力的科学，而含义层次讨论的信息没有形和力，数也只有具体的，没有抽象的，数学的方法只适用于信息载体和外壳这个必要但不是核心的构成部分。显性信息结构是以含义为对象的，所以它的完备需要的是大量的，经常以亿为单位的相关认知事件来实现，而不是寻找一个完美的逻辑推理过程，以小博大。

到现在为止的讨论，存在一个悖论，即显性信息机构完备的标准是当代的认知水平，那么当代认知没有达到的客观存在如何处置，这需要基于显性信息结构的新的认知主体来完成，是另一本书的范畴了。

生命进化的主线索是基因突变、自然选择，但认知革命发生之后，基因突变和自然选择两个原则发生了重大突破，基因突变增加了实验室发生，自然选择增加了人为选择，进化路径改变了。无独有偶，当关于物质、自然界、生命、社会等一系列重要的客观存在对应的显性信息结构达到当代认知水平之后，基于人的认知发展路径也会改变，成为人和机器共同推进认知发展的过程，这个过程不同于过去我们使用计算机或其他工具提升人的认知能力，而是并行发展的认知体系。

第 6 章

信息空间

一个跨越物质与生命，一个经由物质与生命而发展的独立空间，
一个将改变物质空间和生命空间发展规律的新空间。

6.1 信息空间

本节定义信息空间，并分别对其构成要素进行分析。

6.1.1 信息空间的构成

信息空间的构成如图 6.1 所示，有五大要素。第一是构成一个具体信息的载体、外壳和含义，第二是自在态、自有态、记录态三态四类信息，第三是信息结构，第四是一个信息集合的存在空间，第五是在信息发展过程中扮演关键转换角色的主体。

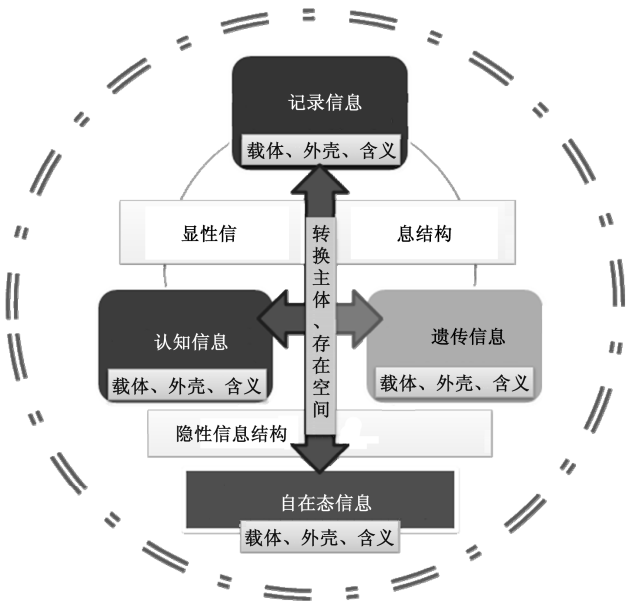


图 6.1 信息空间构件示意图

构成信息空间的首先是一个个分属不同形态、具体的信息。这样的信息是客观存在，作为载体和外壳，有质量、有形态，作为含义，没有质量、没有能量、没有固定的形态。第1章中已经对信息的一些重要属性进行了讨论，而与其他客观存在区别最大的就是由载体、外壳和含义构成一个具体的信息。不管是一个具体的信息三种构成要素都具备，还是载体和外壳合一，如在计算机中处理的信息、生物体的遗传信息，其外壳与载体合一，而只有两种要素，但作为独立存在的信息、作为信息空间的构件，需要同时具备物理形态和含义。载体和外壳的区分不在于其物质构成，而在于是不是主体解释为含义的对象。一个客观存在如果含义不是其存在的主要原因不能称之为信息，不能成为信息空间的构建；没有物理形态则不能称之为客观存在，也不能成为信息空间的构件。

每一个具体的信息都可以归为自在态、自有态、记录态这三态四类信息中的一个类型。所以，从宏观看，信息空间就是由这三态四类构成，信息结构存在于这三态中。如同第2章所讨论的，自在态信息与宇宙同在，与万物共生，所以是无限的。无限不等于包罗万象，自在态信息是客观存在的反映，而不是客观存在本身，这是区分的主要标记。一个记录信息的集合诸如图书馆、一个硬盘上的数据库、一个生物体的细胞、一个人脑，作为整体，只是客观存在，但不是信息。其中的一个个具有独立含义的单体式信息，如一段基因、一本书或一篇文章、一组数据、若干个神经元中包含的内容。作为客观存在，也是信息，但此信息是指作为一个整体所包含的信息，如图书馆的建筑、结构、布局、功能，以及员工数量、书架数量类型、工具实施等，而不是一本本书中的信息。

具体的信息之间存在基于含义的逻辑结构。第3章已经讨论了隐性信息结构和显性信息结构。隐性和显性信息结构都服务于信息的处理和利用，但隐性结构是私有的，服务于拥有该信息结构的主体，显性结构是共有的，服务于一个信息域或主题域的主体。隐性和显性信息结构都是事实和功能兼具的独立存在，不仅包含着信息，还能在条件满足的前提下，实现与该信息所反映的客观存在的功能。隐性信息结构的存在形态与其所代表的具体信息相同，遗传信息和神经单元是自有态信息，记录信息内含的结构是记录态信息。

显性信息结构是共有的，所以需要在已有的信息集合中经过专门的显性工作。目录、索引、知识表示等是记录信息集合显性结构的雏形，目的相同，但还不符合显性信息结构的功能需求。

转换主体和存在空间不是信息，但是信息发展和存在的必要条件，是信息空间的重要构成部分。

这里转换主体是所有对信息进行转换、处理（感知、存储、计算、连接、传输等各项功能）的主体，主体这一概念是泛指，既包括生物体，也包括非生物体。其实转换也是一种处理，但在信息发展中具有独特的作用，所以没有一并归入处理。三态转换是信息发展的关键，从自在态信息到遗传信息、从遗传信息到认知信息、从认知信息到记录信息、从记录信息到显性信息结构，每一次转换都是艰难而又漫长的过程。承担转换功能的主体不断增加，从开始的低等生命体，随着生物体的丰富多样而持续增加，直至智人的诞生；记录信息诞生之后，处理信息的工具不断增加，从机械的到自动的、智能的，并逐渐从计算、传输向感知、认知发展，正在形成与人的功能想媲美的完整处理链环。

这里的存在空间是指一个信息集合占有的或放置的空间。存在空间虽然不属于信息的范畴，但是信息存在和发展的必要条件。没有生物体，遗传和认知信息尽管有载体，但不能发挥其作用。

6.1.2 信息发展阶段及信息空间构成的变化

信息空间在信息发展的每一阶段，其构成和特征有很大的不同。图 4.1 的信息发展阶段，揭示了这一信息空间构成演进的规律。

第一阶段，信息空间只有自在态信息，无边无际、无穷无尽，但是只是物质世界的附庸，消失在时间的长河中，默默地等待生命的诞生。

第二阶段，信息空间增加了遗传信息。我们无法确定生命诞生后多长时间产生了遗传信息，但应该在生命的早期就产生了，因为没有遗传信息，生命的营养系统也不能进化，生命也就无法演进。这个阶段长达 30 多亿年，随生命的发展而发展。遗传信息为信息空间注入了活力，自在态信息经由生物

体与遗传信息相互作用，不断拓展着生物体的信息功能，信息空间中遗传信息的数量在不断增加。数量的增加源自生物体数量增加、生物体类型的增加和一些类型生物体的单体遗传信息量增加。其中对信息增长影响最大的是生物体进化过程中感知和认知功能的生长，为信息的持续进化走出了第一步。这个阶段的信息完全依赖生物体，随生物体的进化而进化，随生物体的存亡而存亡，这个阶段的信息空间是物质空间和生命世界的附庸，解释这个阶段的地球，完全可以不用信息或信息空间这个概念。

第三阶段，发生于约3亿年前，生物体感知和认知功能的发展认知信息在当时的生物体上产生并发展，神经系统的进化在其中起了关键的作用。认知信息提升了生物体的信息能力，也提升了生物体生存能力，自然选择有利于信息能力强，特别是认知能力强的物种，这种持续的正向选择，使生物体认知能力跨上新的台阶，产生了群体间可交流的有声或无声语言。在前期，也是第三阶段的绝大部分时间，认知信息只存在于生物体个体中，不能通过物质空间向同伴发出信息或相互间交流信息，到这个阶段的后期，一些生物体通过长期的积累，可以通过声音或肢体动作发出信息，第一次在地球上出现了脱离生物体的为信息目的的客观存在。这个阶段的信息空间由自在态信息、遗传信息、认知信息和基于认知信息的语言或肢体性外在信息构成，脱离生物体的外在信息为信息空间的下一步发展打开了大门。作为信息空间存在和发展的必要条件，转换主体和存在空间仍然是生物体，这个阶段的外在信息没有保存的条件。

第四阶段，信息空间中的变量来自独立于主体的记录态信息，呈现持续加速增长态势。记录态信息有两种不同的类型，一种是经过人或其他主体认知能力转化过的认知结果，另一种是直接将在自在态信息转换为记录态信息。从这个角度看，感知功能在第二、第三阶段是遗传信息和认知信息发生发展的必要条件，不能直接转换为遗传信息，需要经过人的认知功能系统转换为认知信息；第四阶段则一分为二，通过两条途径转换为记录态信息，也为独立存在的信息空间提供了源源不断的信息客体。因此，从进化看，记录信息是第三阶段后期认知信息发展的自然延伸，从信息空间看，这是一次质变。尽管第三阶段后期产生了脱离生物体的可以短暂独立存在的信息，但存在时

间太短，不能保存，不能摆脱时空的限制。记录信息既是认知的结果，更是提升认知能力的要素。感知和认知能力的提升，记录信息在信息的发展和信息空间发展中扮演了特别重要的角色。由此，第四阶段的信息空间由自在、遗传、认知和记录信息构成，转换主体增加了非生物体的各类传感器和感知、转换系统，为自在的信息空间形成迈出了一大步。

第五阶段，显性信息结构的产生和完善。记录态信息形成了独立的信息空间，显性信息结构这个新的构件则成为信息空间自我发展的基础。本书的第三、四、五三章对显性信息结构的不同侧面做了比较详细的阐述，这里不再重复。作为信息空间的新成员，它有三个重要的功能，首先是记录态信息的抽象，为主体快速、完整、正确地理解其所需要了解领域的信息及其反映的客观存在提供了基础；其次是三态信息在含义层次的中介，实现了三态信息以显性信息结构为基础的统一；第三是具备了重现客观存在及其功能的基本条件，成为走向第六阶段的基础。这个阶段的信息空间从形态上看没有变化，因为显性信息结构本身也是记录态信息，转换主体及存在空间也没有变化，但信息空间的功能变化了，或者说信息空间进化到了一个新的阶段，这个阶段的信息空间具备了不依赖生物体、从自在态信息转换到认知和记录态信息的自我进化能力。

第六阶段，是第五阶段的升级版，信息空间自我的发展和生长可以不依赖生物体，而且已经超越了生物体个体或整体在信息增长中的作用。第六阶段作为客观存在的信息空间构成与第五阶段相同，但功能进一步增强，成为不可逆转的、独立的、自我完善的信息空间。

6.1.3 主体的演变

上面已经提到，作为信息发展和增长的一系列主体不是信息，但是信息空间存在和发展的必要条件，也是分析 6.2，6.3 节的认知进化和信息空间与物理空间、生命空间关系的关键环节，这里做进一步的分析。

在信息发展的第一阶段，没有转换主体，因此信息空间是沉寂的、无为的。第二阶段，产生了生物体感知并转换、生成了遗传信息。第一株生成遗

转基因的古菌，如同在沉寂的信息空间投下了一颗石子，形成了圈圈涟漪，成为信息空间巨变的导火索。从信息空间发展的历程看，这根导火索，直接的结果是遗传信息的不断增长，间接的结果是为信息的第二个转换功能创造条件。

在信息发展的第三阶段，转换主体依然是生物体，但只是具有中枢神经系统的生物体，而且转换的结果变了，将感知的信息转换为认知信息。具备认知信息转换功能的主体从数量上看，大大少于遗传信息转换的主体。到今天为止，地球上生物体数量最多的依然是没有中枢神经系统的生物，尤其是细菌，在数量上占据绝对优势，但转换结果的变化再一次为信息空间增加了重要的新变量。认知信息使信息空间由基本静止发展为动态的，因时因环境而变化的信息。遗传信息使信息空间在变化，但这个变化相对呆板。复制、转录生成一份新的遗传基因，只是增加了量，没有含义意义上的增长；在复制、转录过程中发生变异，但变异的概率比较小，在自然选择后保留下来的更少，变化率很慢。如果将40亿年作分母，所有不同的遗传信息作分子，就可以得出变化率很慢的结论。我们不知道第一次生物体因认知而对外部环境直接做出反应的确切年代和确切行为，但不妨碍我们推断在一个特定的具体和场景，地球上第一次有了因为信息而产生的活动，生命空间与物理空间的互动第一次有了信息的参与，而且这种参与导致了自然选择对认知能力强的物种的倾斜，认知能力由此不断提升，认知信息由此不断丰富，信息空间因扔进第一颗石子引发的涟漪产生了持续激荡的新动力。

第四阶段的转换主体产生了质变，非生物体参与到信息的转换（或前文提到的广义处理）中，而且参与的类型和数量越来越多，在信息发展中的作用越来越大。这里定义为记录信息从感知到传输、存储、处理全过程的工具为处理工具，以区分人类。人在第四阶段依然扮演记录信息形成的主要主体，非生物主体形成的新认知只是辅助；在第四阶段的前期依然是记录信息处理和利用的主要主体，非生物体处于从属地位，到第四阶段后期，信息转换和处理主要由处理工具承担。从以人为主到以处理工具为主的转变过程只有不到一百年，从信息空间发展的漫长历程，或者与以前几个重大转变经历的时间看，相差2~6个数量级。这一转变的直接原因是信息技术革命，而背后的

本质是感知客观存在的需求和记录信息处理利用的需求不断增加，人的能力已经不能满足需求。更重要的是，不仅是从数量上看能力不足，而且是一些感知和处理要求人没有能力满足。需求和能力的差距导致了信息革命的发生，看起来是人类科技的进步，但从历史大尺度看，则与自然选择具有相同的规律，是技术变异，自然选择，适者生存。与以前所有转换主体的演进相比，这个过程变化率很快。几十年的时间，信息传输几乎完全依靠处理工具，大规模的信息处理也已经完全依赖处理工具。从新产生的记录态信息数量看，人直接产生的已经少于由处理工具产生的，而且这个比重将继续快速向处理工具倾斜。强调的一点是，处理工具已经或即将具备科学、技术、管理、艺术类原始记录信息产生的功能，在不太长的时期内，这类信息的生成主体也将以处理工具为主。第四阶段是信息空间走向独立空间的时期，其标志是记录信息，其发展速度的加快源于新的信息转换和处理主体的加入，并为信息空间进入下一个阶段创造了条件。

第五和第六阶段，转换主体类型与第四阶段一样，但能力发生了变化，推动信息和信息空间发展的主要力量逐步从人转向非生物主体。新的能力就是显性信息结构的构建能力，及由此带来的新增认知能力。如同记录信息的形成一样，在显性信息结构构建的初期，是人与处理工具共同完成，以人为主，随着经验、知识的积累和平台功能的完善，工作量以处理工具为主走向全部工作由处理工具为主，人主要承担参与交互、做出判断的角色，并在一定的时候退出整个显性信息结构的构建工作，而这个时间点就是进入信息发展和信息空间发展第六阶段的起点。

6.1.4 信息空间的存在、消亡与扩展

6.1.1 节、6.1.2 节和 6.1.3 节讨论了一般性的信息空间，及其发展阶段，但信息空间不能将基于地球的物理空间和生命空间作为参照系，需要按照信息空间的特点来定义什么是独立的信息空间，它的消亡或扩展。

定义 6.1: 独立信息空间。一个独立信息空间是指空间内存在一个或多个非自在态信息的信息集合，并且与其他信息空间不存在含义之间的有效交流。

这里的有效交流是指对其他信息空间的增长具有实质意义的交流。

推论 6.1: 独立信息空间的消亡。一个独立信息空间消亡是指该信息空间自身不再发展, 它所拥有的信息集合也不能为其他信息空间所利用。

在讨论抽象的信息空间时, 自在态信息作为构成部分, 而讨论具体的信息空间时, 只指自有态和记录态的三类信息。因为自在态信息不独立于物理空间和生命空间, 不具备自我发展的能力。

第二阶段的信息空间是大量的独立信息空间, 一个生物体就是一个独立的信息空间。生命的遗传不是利用了父代遗传基因, 而是父代复制了一份新的成为新生命的遗传基因, 不管父代生产了多少子代, 父代的死亡, 表示该生物体中的信息空间也随之消亡。

第三阶段前期, 在语言或类似语言的肢体信息没有形成之前, 与第一阶段一样, 一个生物体构成一个独立的信息空间, 尽管增加了认知信息, 但该认知信息只能一个生物体自己利用。同样, 一个生物体的死亡意味着以该生物体为存在依据的独立信息空间消亡。到第三阶段后期, 语言或类似语言的超越个体的信息产生之后, 独立信息空间开始小范围扩展, 而延续的时间, 或定义消亡更需要慎重。独立信息空间扩展到一个共同生活的生物体群体, 延续时间需要两个判定条件, 一是是否存在代际增长式传递能力, 二是这个群体的消亡或延续。对于第一个条件, 显然有些生物体没有代际增长式传递能力, 如蚂蚁、蜜蜂、猩猩及很多具有肢体语言, 并以肢体语言告知环境情况或传授生活技能, 但不能积累性代际传递, 一般而言, 除了人类的群体之外, 独立信息空间只增加了主体的数量和小幅度的时间延长。对于第二个条件, 可以判断对于绝大部分的已经拥有语言的原始人类群体, 都能延续若干代而最终因群体的消亡而消亡, 但也有很少的群体, 一直延续到文字的产生, 并形成记录态信息而延续至今。当记录态信息产生之后, 哪怕这个群体消亡了, 但其记录下来的信息会融入到新的信息空间中。

到第四阶段, 记录态信息成为信息空间增长的主要形式, 独立信息空间的时间延续基本上一直延续到现在并将继续延续下去。从空间看, 独立信息空间覆盖的地域日益扩大, 一直到整个地球文明成为一个独立的信息空间。记录信息是独立信息空间扩展、整合的主要原因, 当然, 记录信息的发展及

语言障碍的突破是以人类认知能力的发展和经济的发展紧密相关的。

6.1.5 信息空间进化的几个规律

以地球为范畴，信息空间的进化区分出六个阶段，第五阶段将要进入，可以看到前进的轨迹，第六阶段还需要一段时间，进入第六阶段后的信息空间或信息发展进程是否如本书的预测，需要等待未来的验证。而超越地球文明的统一信息空间，更是展望而已。因此，全面总结信息空间的进化规律为时尚早，但从过去的四个阶段，可以看到几个基本的规律。

首先是关于独立信息空间增长速度的规律。

定义 6.2: 延续协同定律。在一个独立信息空间中，转换主体的数量、延续的时长、协同平台的质量是该信息空间规模增长和进化速度决定了该空间增长的速度，称之为延续协同定律。

这个定律可以表述为：

$$\hat{U} = \Gamma \times \omega \times \rho \quad (6.1)$$

其中 \hat{U} 为一个信息空间的增长速度； Γ 为主体的数量，包括认知和处理工具两类主体，值域为 $0 \sim n$ ，以人为单位，处理工具按处理能力折算为人数； ω 为延续的时长，值域为 $1 \sim m$ ，以独立信息空间生物体一个生命周期为单位； ρ 为协同平台质量，指平台的性能和平台上参与协同的信息的质量和主体的认知水平和能力，值域是为 1 到 100，或其他适当的整数值，是一个具有放大作用的系数。如果一个独立空间存在一个主体，延续时长为一个生命周期，那么它的增长速度为 1。

在信息空间的第一阶段，主体数量是零，所以尽管无限的自在态信息无处不在，信息空间没有增长。第二阶段，独立信息空间只有一个主体，延续时长就是这个主体的生命周期，也没有协同平台，每个独立信息空间已几近于零的速度增长，但主体的数量巨大，生命周期虽短，但每个生命周期都可能产生基因变异，这是这种大数量和快频率，用了约 40 亿年时间发展到具有最高水平的认知生物体——智人。在第三阶段后期，第一次在主体数量和延续时长都有了一个小的增加，平台这个系数依然为 1，一个独立信息空间的增

长速度提升了一到两个数量级。

进入第四阶段，独立信息空间中三个变量都开始正向增加。首先是主体扩展到了记录信息扩展到的人群中，而这个数字每过一个时期就有大的增长，一直增长到今天地球上能为记录信息增长做出贡献的几十亿人。其次是延续时长自从文字和可保存的记录信息出现之后，就与地球时间同步，地球上过了多少年代，折合人的平均寿命多少代，延续时长就是多少。第三是协同平台在这个阶段的后期诞生，并为协同能力的提升做出了日益增加的贡献，这个系数从社会性的信息交流渠道诞生开始就大于 1，随着信息交流渠道的增加、技术手段发展、效率和参与人数的增加而持续提升，互联网全球普及使这个系数突破 50，并向 100 进发。第五和第六阶段还没有开始，但可以预见，这个规律将继续发挥作用。

这个定律所揭示的延续就是跨态的延续，是不同的生物与/或物理现象在一只看不见的手指引导下，实现了信息空间延续发展的规律。这个定律揭示的协同就是信息空间进化从串行走向并行，从小规模并行到大规模并行，进化速度呈指数增长的规律。并行实现了指数增长速度，但有条件，那就是平台是否能够有效地积累并行的认知增长成果，主体能否有效地利用这一成果。

其次是关于信息空间跃迁式发展的规律。

定义 6.3: 积累跃迁定律。信息空间的发展呈台阶式跃迁，每一个台阶呈现长期渐进积累后跃迁到一个新台阶规律，称之为积累跃迁定律。

在信息空间发展的每一个阶段都可以看到遵循这一规律。自在态信息对无数原始生命超过 10 亿年的激励和选择，跨越到信息空间包含遗传信息的发展阶段；具有遗传基因的生命再经过超过十亿年的变异和选择，信息空间加入了认知信息这个新成员，跨越第二台阶；具备认知系统的生命体再次经历 5 亿年左右的积累，终于突破生物体的屏障，产生了以文字为基础的记录信息，信息空间实现一次最重大的跃迁，以信息为目的的客观存在确立了信息空间的独立性；在这个中间，从语言这种外在信息并使文字产生的直接催化剂的产生到真正以文字为基础的记录态信息产生，又是经过百万年计的进化过程。文字性记录的产生到记录态信息快速增长，超越人的利用能力的发展还是一个长期积累，最近数十年，特别是最近几年呈现需要通过一个出发点，进入

显性信息结构的发展阶段。这些大阶段、小阶段的积累跃迁模式，表明每一个阶段都呈现如图 6.2 所示的形态。

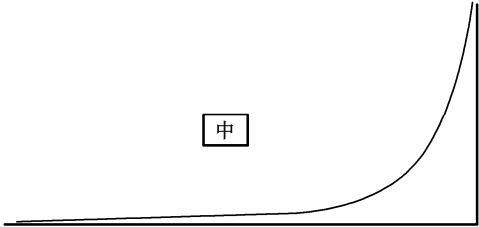


图 6.2 积累跃迁模式示意图

其中横轴为时间，纵轴是数量。

按照这一规律，以及目前可见的实践轨迹，信息空间从第五阶段发展到第六阶段只需要几个十年。

第三是信息空间构成中非组合规律。

定义 6.4：非组合规律。非组合规律是指一个信息空间的组成部分各自独立，不能通过几个或一定数量的基本构件组合成一个个具体的构成部分。换言之，信息空间中构件的增长，不能同物质或生命一样几个基本粒子的不同组合，决定着物质的构成和性质；几个核苷酸的不同组合决定着遗传基因的遗传性状，即使是在不同的构件中存在相同的部件，也不能约简。一对一地反映相应的客观存在，一对一地在隐形或显性信息结构中体现，一对一地复制、处理。在一定的场景下，为减少冗余，通过指代等方式约简重复性描述，但对于信息单元或信息结构本身，却必须与客观存在一一对应。这正是信息空间的构成特征，也是信息空间的增长特征。

6.2 信息空间与认知能力发展的关系

信息空间的发展与认知能力的提升息息相关，信息空间发展是认知能力提升的前提和基础。

6.2.1 认知能力与信息空间发展的一般关系

这里我们仅讨论与信息空间发展相关的认知能力，而不是从认知科学的角度讨论。

如图 6.3 所示，相对于一个独立信息空间的认知能力有三个要素：主体、信息和技术。这里主体是指具有认知能力的人（或其他具备认知能力的生命体）或系统（非生命体），信息指这个独立信息空间的主体所能利用的信息，技术包括与主体获取信息、利用信息相关的所有技术或基础设施。

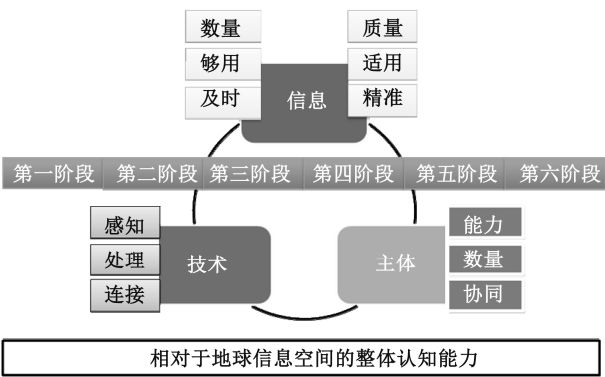


图 6.3 认知能力与信息空间发展一般关系分析示意图

在本节，我们把信息空间发展的第一到第三阶段称为信息空间发展的前期，第四和第五阶段为中期，第六阶段为后期，并按这个分期逐一讨论，同时，还讨论假定信息是微观物质世界构成的要素，对认知发展的影响。在这里，我们将信息空间范围定义在地球这个物理空间内。对应于整体认知能力和信息空间的发展阶段，可以归纳出如表 6.1 所示的相互关系。

表 6.1 整体认知能力与信息空间发展阶段的关系

要素构成				前期	中期	后期
主 体	能 力	个 体	人	逐步形成	逐步增强、增速趋缓	增长停滞
			生物体	逐步形成	增长停滞	增长停滞
		群 体	人	无—弱	参与数量程度逐步提升	增长停滞
			生物体	无—弱	增长停滞	增长停滞
		非生物体		无	弱	逐步提升
	数 量	人		从 0~百万	从百万到数十亿	数十亿
		其他生物体		1×n	1×n	1×n
		非生物体		无	少	逐步增加，总能力超越生物体到主要认知主体
	协 同	自发		无—弱	随外部环境发展逐步增强	成为必要条件
		平台		无	数量逐步增加，质量逐步提升	成为基础设施
信 息	数 量	够用		被动	逐步增加，不足与过多并存	从过多到个性定量
		及时		被动	逐步提升，大部分不能满足	从大部分不满足到基本满足
	质 量	适用		被动	逐步提升，满足的只有少数	从少数满足到基本满足
		精准		被动	逐步提升，满足的数量更少	从很少满足到基本满足
技 术	人	感知		无	无—弱—强	增速趋缓—停滞
		连接		无	无—弱—强，泛在、实时	增速趋缓—停滞
		处理		无	弱—强	增速趋缓—停滞
技 术	非生物体	感知		无	无—弱—强	成为该空间主要的感知来源
		连接		无	无—弱—强	成为该空间主要的连接构成
		处理		无	无—弱—强	成为该空间主要的处理主体

下面将对这三个阶段所对应的地球信息空间认知能力进行初步的分析。

6.2.2 信息空间发展初期的认知能力

第一阶段,信息空间没有转换主体,是一个沉寂的、漫无边际的自在态信息在无穷无尽的时空中流淌。就信息而言,没有反应、没有进化,默默地等待着一个时机。这个时机在地球上第一个生命诞生的那一天终于到来,信息空间发展进入第二阶段,产生了生物体感知并转换,并经历漫长的偶然性与必然性交织的发展过程,产生了遗传信息。更重要的是,无论在这几十亿年时间内发生了什么变化,提升认知能力的遗传基因总是沿着正向进化发展。这个时期信息空间的认知能力集中在生物体认知能力遗传基因进化,可称为生物体认知功能形成前期。在信息空间发展的第三阶段,转换主体依然是生物体,但生物体已经形成了中枢神经系统,并不断朝着功能完善的方向演进。随着中枢神经系统的进化,生物体对自在态信息的转换进入新阶段,将感知的信息转换为认知信息。认知信息的产生对生物体认知能力的提升产生了双重正向演进激励:中枢神经系统的记忆功能和走向群体认知进化的表达能力。地球环境的变化及自然竞争适者生存法则,导致了一些生物体产生了记忆和群体间交流功能,蚂蚁、蜜蜂、猩猩等大量生物物种具备这种功能,并持续向具有抽象概念的语言进化,直到十万年前智人的产生,记忆和语言功能最终在遗传基因层面成熟。再经过九万多年的发展,记忆和语言功能得到长足的进步,为文字的产生完成了准备。信息空间发展的第三阶段是认知功能形成后期及群体认知能力进化前期。随着认知能力不断前行,能跟上这个步伐的生物体种类越来越少,具有简单认知信息形成功能的生物种类只是很少一部分,而发展到具有相当复杂程度的抽象概念作为语言基础的就只有人类了。

将上述信息空间发展的前三个阶段与认知功能发展的关系做一个总结,在整体上是信息空间认知能力的多中心蓄积期。之所以称为多中心,是因为在地球上,在不同的物种间、不同的智人群体甚至大量的不同个体间各自发展,没有形成比较有效的群体进化能力。这个蓄积期是信息空间发展中基础的、不可越过的、最为艰险的、历时最长的、又是从认知的角度看最难与信息空间的发展连接在一起的特殊时期。

从以信息为基础的认知能力发展阶段看,到第三阶段后期,又进入一个

瓶颈，那就是语言这种信息表达和交流方式受到个体或小群体的认知能力和时间、空间约束，如果没有一种新的方式，认知能力就会就此停滞，如同数百万年到如今的一些动物的认知进程一样，停留在智人水平。

6.2.3 信息空间发展中期的认知能力

这个突破就是能脱离人的记忆依然存在的记录信息。记录信息的产生和发展是信息空间发展到第四阶段的标志，也是以信息为基础的认知能力的一次飞跃。这是一个从潜在、蓄积走向显现、新客体产生并快速发展的阶段，但所有的发展均依赖于人的认知能力，所以从以信息为基础的认知能力发展看是一个被动发展的阶段。

以五千年前比较成系统的记录产生为标记，第四阶段可以还分成前后两期。在第四阶段前期，记录信息的产生和发展依然十分缓慢，但比认知信息及语言的产生要快得多，所经历的风险也小得多。从大量人类群居部落以结绳记事、各种器皿或物体上刻画的早期准文字，从四大古文明在相距不长的时间内各自独立形成文字的历史看，人类从语言进化到文字是一个必然的过程。到后期，如果以工业革命和信息革命结合点的自动化系统产生为第四阶段和第五阶段的分界线，则经历了不到五千年的过程。这个过程的特征是社会记录信息的能力不断快速增长，增长速度远远快于同期经济社会或生活水平提升的速度。记录的主体从少数部落的少数几个人扩展到数十亿人，记录方式从刀刻、手写不同技术水平的印刷。更重要的是，记录信息的产生对人类认知能力的提升起到了决定性的作用。实际上，从智人诞生之后十万年间，从遗传基因的角度看，认知能力几乎没有发生重大变化，但认知能力却不断提高。以语言为起点，记录信息产生为加速器，人类认知了越来越多的未知世界，生存能力不断提升，认知的起点越来越高。特别是文艺复兴之后的科学大发现及随后的工业革命，认知能力变成了社会生产力。这种生产力的提升进一步加速了具有记录能力、研究能力的总人数的增长，加快了记录信息的设备的发明和增长。记录信息的增长，独立存在的信息空间逐渐形成规模。

然而,这种增长和发展态势产生了两个需要解决的问题,一是在很多领域,人的记录能力不能满足研究和生产、生活、管理的需要;二是记录信息的数量持续增长,对人类自身的利用能力带来了挑战,需要新的技术、方法和工具来辅助提升利用记录信息的效率和质量。这种需求推动信息空间进入第五阶段。

提升信息空间发展的第五阶段对认知能力的技术、方法、工具主要由以下几方面构成:感知、结构化、连接、处理和信息。感知的需求首先产生于人类对探索未知的本能和生产管理过程自动化的需要。对于探索未知,不管是太空、生命还是物质结构,一系列新的科学仪器设备提升了人类的感知能力,也提升了自在态信息的感知和记录能力,提升了代表人类认知水平记录信息的发展能力。对于生产和管理,利用人类自身的感知能力既存在成本过高的问题,更有超越人类能力极限的问题,因此满足生产和管理的感知器随着经济社会的发展类型不断增加、能力不断提升,直接与计算机信息系统连接的生产和管理感知设备,在信息空间的发展历史上第一次产生了不经过人类的记录信息,这是有一次质的飞跃,为信息空间的发展从被动走向主动铺下了第一块砖。当第一个通过传感技术获取信息实现自动控制的系统成功之后,人类社会一发不可收拾,几十年的时间内,发展到无处不在,记录的信息量超过经由人类自己生产的信息量。

对信息集合结构化既是对记录信息不断增长,如何提升人类认知能力跟不上的矛盾,也是提高信息相对于应用需求质量的关键环节。记录信息集合结构化的努力至少可以追溯到数千年以前,几个古文明发源地统治者收集的记录信息量已经很庞大,需要专门的整理加工,分门别类,为当时的知识分子或政府机构、官员利用达到按图索骥的作用,记录信息集合结构化逐渐形成了稳定的方法体系,分类、主题、索引成为大小记录信息集合结构化的普遍工具。而这种努力在计算机产生之后,进入了新的发展阶段,一方面是信息技术的加入使记录信息的增长和处理达到了一个新的高度,记录信息数量和处理能力激增对结构化的要求日趋强烈。另一方面是人类从几百年前设想发明一种计算机开始,就不懈地追求如何替代人的智能。计算机发明后不久,人工智能系统就在这种背景下粉墨登场,而人工智能的每一个进步都与

信息的结构化联系在一起。在 6.2.4 节可以看到，记录信息的结构化是信息空间从被动发展到主动发展的关键要素。

记录信息的产生超越了时空的约束，连接就成为信息有效利用的必要技术因素，也成为信息空间存在和发展的重要构成要素。烽火、鼓声、驿站、邮政系统都是信息传递的渠道，因此也是信息连接的一种方式，这是第四阶段的技术手段。到了第五阶段，信息空间连接的主渠道是基于信息技术的信息网络。信息网络这一连接模式第一次利用了信息含义、外壳、载体可分离、可重组的特点。例如，传送一本书的信息，不再需要将物质性的书从甲地送到乙地，经过转换，通过电子方式就可以实现。这种模式的连接在信息空间发展的第五阶段，不管是对人类还是电子信息系统，都经历了从无到有、从弱到强的发展阶段。连接的发展直接扩大了独立信息空间的范畴，使人类利用信息逐步从比较狭小的区域发展到全球。如果不计语言 and 知识领域的约束，到互联网在全球普及的时候，全球独立信息空间已经连接为一个。所以对于信息空间认知能力的提升，连接的功能就是将进化由串行模式切换到并行，认知能力的进化不再是一个主体个人的能力，而是在个体能力的基础上利用了整个信息空间的认知主体的力量，这也是信息空间特殊的发展规律的一个基石。

记录信息的持续增长以及信息利用时间约束，提出了越来越高、越来越复杂的信息处理需求，基于计算机的信息处理，成为第五阶段信息空间发展的主要特点之一。如果不计像算盘、对数计算尺、打卡机等机械的、辅助的记录信息处理工具，基于信息技术的记录信息处理在第五阶段同样经历了从无到有、从弱到强的过程。到 20 世纪末，除开科学研究、艺术创作等少数领域，计算机信息系统信息处理能力和承担的处理任务远远超过人的能力和人承担的任务，不需要人干预的计算机信息系统信息处理领域在数量上将逐渐超过仍然依赖人的信息处理领域。也就是说，在第五阶段的后期，记录信息的增长、信息空间的发展，技术力量以成为主要因素。

记录信息在不断增长，连接和处理的技术也在不断发展，但从应用的角度看，信息的数量和质量在总体上还不能很好地满足需求。从记录信息数量的两个指标，信息的够用性出现两种不同的倾向，一是对部分应用出现了过

多的现象，二是对另一部分应用依然是不足。对信息数量的第二个指标，及时性，同样是在逐步提升，但还达不到相当部分需求的及时性要求。产生的原因首先是信息获取能力、感知能力和连接能力的持续提升，使互联网上容纳了越来越多的信息，而任何个体或系统在给定时间内能够处理的信息量总是有限的，互联网上拥有的相关信息量超过了处理或连接能力，产生过多或不及现象。其次是信息发布、传播的渠道日益丰富，各类专业期刊杂志登载的信息数量巨大，不少领域非互联网的信息也超越了人的利用能力。第三是对网上网下质量不一、重复众多的记录信息缺乏有效的工具和机制，进行比较全面的结构化处理导致过载或不及时；第四是没有一种机制，为信息不足的应用提供全面、长远的解决方案，导致信息数量和质量不能满足需求。第五就是记录信息结构化处于自发应对记录信息增长和追求人工智能发展的副产品，没有理论基础和基于理论的方法和工具，成为信息利用数量和质量、认知能力提升的主要障碍。

综合信息空间发展中期两个阶段，信息空间成为独立空间，而且空间构成不断丰富、扩展，基于信息空间的认知能力整体处于被动发展期，但为进入主动发展期在各个方面打下了基础。

6.2.4 信息空间发展后期的认知能力

信息空间发展的第六阶段还没有到来，进入这个阶段还需要一段时间，由于这几年记录信息的高速增长（大数据）和各个领域信息技术应用趋向智能化的趋势，在加速这个阶段的到来。

推动信息空间的发展从中期进入后期的主要动力是人类的认知能力增长逐渐趋缓并趋于停滞。判断是否进入信息空间发展后期的标志是从被动增长发展到主动增长，相对于一个信息空间的认知能力提升从以人为主转向以非生物智能体为主。这一过程实现的主要基础是信息空间驱动的记录信息结构显性及其完备，并以逐步完备的显性信息结构驱动各种以信息技术为基础的设备、设施，提升对未知探索的能力，管理和操纵社会的各类活动，经济的、社会的、生活的、军事的等。

进入这个阶段，以非生物体为主体的认知能力摆脱了人类的生理约束，在感知、连接、处理能力达到新的高度，信息的数量和质量逐步满足整个社会各种需求，并为地球文明走向太空奠定了基础。这也是本节将主题定格在信息空间的认知能力发展，而不是信息空间本身发展的原因。信息空间形成独立的认知能力，克服基于生命的认知能力发展障碍是信息空间发展的方向，也是必然趋势。

6.2.5 一个假设：如果信息是物理世界的构成要素

到目前为止，本书的信息只包括了三态四类，没有分析信息是否是物理世界构成的必要组成部分，如果是的话，信息空间构成将增加新的类型，发展的历程也将重写。

这里所说的信息是物理世界的构成要素不是生物信息学或量子计算、量子通信及量子观测中量子信息衰减所讨论的信息。“生物信息学的任务是建立和使用生物学数据库，其中一些包含基因组的 DNA 序列。生物信息学对于从庞大的生物学数据中挖掘有关基因结构和表达的有用知识而言是必需的”，¹“本书在简要介绍分子生物学和化学后，详细分析了各种算法，从不同角度展示计算机为分析海量并且非常复杂的数据提供难以估计的帮助”。²量子计算和量子通信则是利用量子运动中态的变化特征作为计算或通信的一种新的技术方法。量子测量过程中的量子信息衰减是指一个量子系统中具体量子状态观测信息的减少直至到零的现象³。这些都是关于量子态的信息，而不是关于量子本身结构的信息。

如何理解作为物理世界构成要素的信息，今天的物理学和分子生物学或生命科学还没有给出明确的答案，但迄今为止的研究成果，至少提供了两条

1 分子生物学（原书第五版），[美]Robert F. Weaver 著，郑用琰、马纪、李玉花、罗杰等译，科学出版社，2013 年，第 849 页。

2 生物信息学概论，[美]奎恩、雷默著，孙啸等译，清华大学出版社，2004 年，前言。

3 量子信息物理原理，张永德著，科学出版社，2006 年，第 6 章。

途径,生物分子及生物体自身携带的与功能相关的结构信息和进化信息、物质携带的关于自身结构的信息。

对于生物分子及生物体自身携带的与功能相关的结构信息和进化信息,基因的转录、翻译、复制、重组、转座等功能的研究及基因组学的研究证明这样的信息是存在的⁴,需要回答的问题是,这类信息是遗传信息还是另一种自有信息。分子生物学和生命科学迄今为止的研究成果,基本上证明这类信息是遗传信息的组成部分,在基因的转录、翻译、复制、重组、转座及其他基因组的功能实现过程中,基因结构或基因组所表达出来的信息视作基因信息是合理的⁵。然而,分子生物学和生命科学还有很多未知领域需要探索,还有很多未知。对于生物体自有信息的研究而言,至少还有五个主题需要进一步探索。第一是没有编码的基因如何、在多大程度上、以什么方式影响生物体遗传过程及生长过程,特别是这种方式在性质上是否与遗传信息相同。第二是编码与/或非编码基因的未知功能结构。尽管确实存在这样的未知功能结构,但从信息类型看,还是属于遗传基因这类自有信息。第三是以蛋白质组学为代表的生物体内非遗传基因的生命过程控制信息⁶,第四是认知信息及功能的发展对遗传基因或其他生命过程控制信息的影响,这两个主题的研究如果揭示出这样的控制过程和影响模式不同于遗传信息,有可能产生除遗传和认知之外的第三类生物体自有信息。第五是细胞本身的信息功能。在过去的生命科学的研究中,没有将细胞的信息功能作为一个单独的、基本的构成要素,而是作为生命过程中某些阶段存在的功能。从信息的角度看,第一个生命细胞的诞生即存在信息功能,在生命生存和进化的全过程中,信息功能一直被视作所有细胞的基本功能。

对于物质携带的关于自身结构的信息,或者说信息是不是物理世界不同结构层次的必要组成部分,目前的研究还不足以做出肯定的答复。根据《量子信息物理原理》一书,量子信息研究只是如何测定量子态及如何利用可以

4 分子生物学(原书第五版),[美]Robert F. Weaver 著,郑用琰、马纪、李玉花、罗杰等译,科学出版社,2013年。

5 同上。

6 同上,第818—849页。

观测的量子态用作计算或通信，没有回答关于量子构成或决定量子态要素中是否有信息的存在。也就是说根据量子电动力学（QED），基本粒子不是具备质量就是携带力，而不存在既没有质量，也没有力的信息存在空间。但在另一个研究领域，如同霍金在《果壳中的宇宙》和《时间简史》两本著作中指出的，物质进入黑洞前，出现虚粒子对并相互湮灭，其中一个落入黑洞，另一个逃逸，并作为波储存在时空另一维的 P 膜上^{7, 8}。一种说法是，逃逸的粒子携带正质量，被黑洞俘获的粒子携带负质量，但为什么不能假设逃逸的粒子没有质量，只带走了信息？也就是说根据量子色动力学（QCD），夸克对构成的虚粒子中有一个可能具有信息的属性。但不管如何解释黑洞信息丢失问题，无论是巨大恒星坍缩为黑洞（黑星），以及黑洞所吞噬的物质或构成黑星的物质内含信息这一结论没有人否定⁹。物理学正在研究这些信息到了哪里，黑洞在霍金辐射发出没有信息的粒子又如何重新获得信息。

如果这个假设成立，信息是物理空间物质的组成部分，而不仅是反映客观存在的含义，则信息又增加了一类自在态信息。分子生物学的研究成果已经证明，信息是生命空间的必要组成部分。这样，三个空间相互依存、相互影响，形成新的宇宙观，这是下一节讨论的主要内容。

6.3 信息空间与物理空间和生命空间的关系

物理空间、生命空间、信息空间三者之间存在着千丝万缕的联系，推动着宇宙演进，本节分别从信息空间与物理空间、信息空间与生命空间、三个空间互动下的宇宙演进三个角度阐述。

7 果壳中的宇宙，[英]史蒂芬·霍金著，吴忠超译，湖南科学技术出版社，2001年，第4章。

8 时间简史，[英]史蒂芬·霍金著，吴忠超译，湖南科学技术出版社，2008年，第一版第十二次印刷，第六章和第七章。

9 莫须有的黑洞，Carlos Barcelo、Stefano Liberati、Sebastiano Sonego and Matt Visser 著，虞骏译，载《宇宙与天体》，电子工业出版社，2012年，第47-53页。

6.3.1 信息空间与物理空间

信息空间与物理空间之间存在着三类,严格地说存在着 $2+1$ 类相互关系:第一类自在态信息是信息空间发展的基础、认知和记录态信息改变着物理世界,第二类自在态信息与信息空间的互动改变物理宇宙的发展轨迹。下面分别讨论。

从本书第1章开始,已经从信息的定义、形态、转换、运动、结构、增长及信息空间发展的各个侧面分析了第一类自在态信息在信息空间构成和发展中的作用,这里不再赘述。

如果从月亮或太阳的角度看地球,从农业革命以来,地球上的植物和植被发生着人为的变化;从工业革命以来,除开植物和植被的变化继续发生并有一定的加速外,人造建筑和人造物品不断增加,人造道路持续增长,能源的利用不断提高,人为因素在地球温度变化中扮演了重要角色,更有一些人造物登陆月球,甚至在太阳系游弋。这是人类认知能力发展的结果,而认知能力与信息空间发展之间互动关系已经在上一节讨论。没有记录信息对认知能力的贡献,不可能产生农业革命和工业革命,更不用说信息革命,当然这一过程是由生命空间一起参与的。信息空间经由认知能力改变物理空间已经大量发生,随着信息空间进入新的阶段,这种变革能力将持续增长,对物理空间现状的改变将更加深入、广泛。

第二类自在态信息目前只是假设。如果说,它只是在视界附近失去了量子态存在的速度、质量、位置信息,那么它还是第一类自在态信息。如果既是信息态,又能成为夸克、介子等基本粒子形成的要素,则是第二类自在态信息。即使第二类自在态信息存在,目前为止,也只是在信息空间的构成中增加了一种新类型,而没有成为信息空间发展演进的新因素,或者说,还没有介入到信息空间的发展进程中。

6.3.2 信息空间与生命空间

如果说物理空间是信息空间之父,则生命空间是信息空间之母。这个结

论在一定的范畴内是正确的，但如果第二类自在态信息的假设成立，信息空间与物理空间的关系改变；如果由于信息空间的发展而重塑遗传基因并由此重塑新的生物体，这样的生物体在地球之外的空间生存发展，这个结论也将调整。

迄今为止的研究成果表明，生命的诞生源自物理空间，但此后生命空间的每一步发展都是与信息空间互动的结果。遗传基因和认知信息是生命空间的构成部分，也是信息空间的构成部分。记录信息源自生物体的认知能力，又独立于生命且反过来影响认知能力的发展。如果没有记录信息的诞生和发展，分子生物学和遗传工程不可能形成，通过遗传工程改变基因，从而主动干预生命空间的局面也不可能存在。

如果存在直接影响生物体生命存续期间行为的第三类生物体自有信息，而且这类信息在信息空间中不是孤立的，而是与记录信息、显性信息结构可以互动的信息，则不仅进一步拉近了信息空间与生命空间的距离，更深化了信息空间与生命空间的互动。

6.3.3 三个空间互动下的宇宙演进

几千年科学探索，似乎我们已经对物理空间和生命空间有足够清晰的了解，一组物理定律决定了物理空间的发生发展，甚至决定了生命空间的发生和发展，但真实情况并非如此。

就物理空间而言，还有大量的现象并不能用这一组定律解释，如占已经观察到的宇宙总质量约 95% 的暗物质和暗能量。物理学家依然试图用这一组定律去观察、去解释，但为什么不能有非物理的视图呢。就生命空间而言，尽管生物体的产生及其作为一个自然界客观存在，遵循这一组物理定律，但生物体的行为及人的思维等在生命之上产生的东西，具有超越物理定律的能力。尽管有很多论据说明生命产生的必然性，难道就可以说明生命只是物理世界的产物，而不是物理世界的改变者，或者说这个宇宙的掘墓者。

显然，很多科学家会对此不屑一顾，认为人怎么可能改变由这一组定律决定的物理空间，更不用说可以在更高层面决定万物运动的 M 理论。数千年

来探知的这一组定律是人类认知的结果，穷尽了人类智慧。由于生命本身遵循这一组定律，生命周期及脑功能极限，使今天和今后的人难以超越迄今为止的科学巨人，因此，仅从生命空间的认知能力推论，人类顶尖的认知能力正在走向停滞。然而，加上在生命空间诞生的信息空间呢？记录信息的诞生拓展了人的认知能力，信息结构的诞生提升了人利用记录信息的能力，但这两个能力已经成为人类认知的基础。具有属性和功能双重能力的显性信息结构及其完备，在早期可以再次提升人类的认知能力，在远期可以形成比人类认知能力更强的能力，从而改变我们观察到的、遵循这一组物理定律的宇宙本身。

我们不仅需要 M 理论去解释千姿百态的浩瀚宇宙¹⁰，更需要我们的认知能力能在这样的宇宙定律（如果存在的话）作用下，使这个宇宙更加丰富多彩，使宇宙的存在及其发展有更多的变量和前景，而不是决定论下的周而复始。这是信息空间诞生后，物理空间、生命空间、信息空间之不同定律相互作用、相互影响的必然结果。

物理世界所追求的是以统一的力的理论去解释包括生命在内的宇宙发生和发展，但是，生命空间只在以物质为基础的行为遵循物理定律，以生命和认知能力为基础的行为不遵循物理定律，其特征就是非物理规律的多样化。同样，信息空间以物质为基础的行为遵循物理定律，以生命为基础的行为遵循生命定律，而独立存在的信息含义的行为不遵循物理定律和生命定律，其特征就是发现、利用并改变既存的物理和生命定律，已经在改变生命空间的发展，为什么一定认为不可能改变物理空间的运动。

按照物理定律，地球已届中年，正在逐渐走向不适合生命存在的天体，太阳几十亿年后也将经历能量耗尽，成为红巨星、白矮星的过程。根据物理定律生存的生命过程，不具备迁移到另一个相对于人类极为遥远、宇宙中只有极小概率存在、适合人类生存的天体上，如此，一个宇宙按物理定律创造的蓝色星球及其多彩的一页就此翻过。

一定如此吗？生命一诞生就开始改变物理存在。相对于地球的体积，如

10 大设计，[英]史蒂芬·霍金著，吴忠超译，湖南科学技术出版社，2001年，第154页。

此渺小的细菌却改变了地球的大气层，使本不适合生命存在的大气构成发生重大调整。此后生命又神奇地分成自养和异养两大类型，为生命的延续提供了必要条件。确实，这样的过程完全遵循物理定律，但物理定律自身不能形成这样的变化。这就是生命空间的行为不是因为物理定律的存在而发生，那么多物理定律，为什么只是遵循了二氧化碳到氧气的转换，而不是别的。如果遵循了另外的物理规律，生命空间就不是今天的样子，也许根本就不存在了。因此说，生命空间的诞生与发展，是生物体创造的，而不是物理空间恩赐的。当生物体认知系统发展到形成意识，产生有意识的行为时，这些行为尽管在于物质相关的过程要遵循物理定律，但行为因何产生不遵循物理定律。遗传基因的转录、翻译、复制过程从表象看是物理、化学过程，遵循物理定律，但四种核苷酸不同排列及具有不同功能的基因组结构形成的遗传信息不是物理定律决定的。更不用说，不是由物理定律决定人际关系、喜怒哀乐、战争与和平、贫富的均衡与不均衡、技术和产业发展的规模与水平。这些行为遵循生命定律，生物演进、基因、认知、经济、政治、文化、军事的规律和准则都是生命定律的局部。

信息空间发生和发展的过程同样在遵循物理定律的同时，还遵循不同于物理定律的生命定律和信息定律。在本书第 5 章已经分析了信息发生发展的规律和准则中遵循物理定律、生命定律以及不遵循这两个定律的原因和例子，这里再做一次简要的系统分析。

信息的发生发展遵循物理是因为信息的载体和外壳是物理的，当信息做不与其载体和外壳分离的运动时，遵循物理规律；而信息做与载体和外壳分离的运动时，含义没有质量、没有力、没有固定形态的基本特征决定了它不遵循物理规律。在人获取一条信息的过程中，不管是看到、听到还是身体感知到，其转换和传输过程都可以用物理定律解释，但这样的化学信号或电信号到达大脑，进入一个先验的信息框架，并据此做出反应或保存在该信息框架时，实际发生作用的不是具体的化学或电信号，而是含义本身。理解这个过程，比理解量子纠缠、量子引力要容易得多，但这个抽象过程不属于物理定律。当信息的运动过程依赖生物体是，遵循相应的生命定律，如果脱离了生物体则不遵循生命定律。生物体感知、传输信息必然依赖生物体的神经系统，生物体处理、保存信息，必然借助认知系统。但一旦脱离了生物体，就

不再遵循生命定律。尽管信息空间的发展看起来是在物理空间和生命空间的哺育下成长,其实如同生命必须诞生在物理空间、依赖物理空间发展,但逐步形成独立于物理空间的发展轨迹和定律一样,信息空间也逐步形成了独立于物理空间和生命空间的信息空间及其独特的发展定律。在自在态信息的基础上经由遗传信息和认知信息发展到记录信息,经由记录信息发展到显性信息结构,经由显性信息结构到形成独立的自我完善、自主发挥作用的信息空间,每一步都充满不定性、经历了漫长的过程,但有理由相信,最后两步,比信息空间发展的过去几个阶段要简单,甚至比形成语言和文字都要简单得多,正在加速成为现实。

从宇宙演进的维度看,物理空间的逻辑是能量、质量的守恒和力的作用和统一,在物理空间多变的现象中寻找不变的规律;生命空间的逻辑是在物理空间逻辑的约束下,在不变中寻找可变的生存和发展;信息空间的逻辑是在物理空间和生命空间逻辑的约束下,寻找如何改变物理空间和生命空间运行轨迹的规律。

假定我们观察到的这个宇宙迄今为止按物理规律演进,但不能说其他宇宙或这个宇宙的未来必然按这样的物理定律进化。一株细菌的诞生,创造了地球生命空间,以今天的人类能力去改变太阳系、改变银河系,直至改变宇宙的演进,难道比细菌创造地球文明更难吗?几十亿年沉寂的信息空间,在有了生命的转换之后,又经过几十亿年才形成了记录信息,但记录信息形成之后几千年时间,显性信息结构开始形成,走向独立的、自我发展的信息空间技术条件基本具备,历史正在经历信息空间走向第六阶段,信息空间的定律正在形成和完善。自主的信息空间将弥补人的生理不足,将认知的触角伸向今天未知的领域。如果暗物质和暗能量不遵循这个宇宙的物理定律,如果我们能发现通过构成物质的信息来利用这些物质和能量,太阳为什么一定要走向红巨星;如果我们能够保存地球的大气层,保护地球的环境,能够从太阳或其他能提供热能的星体持续获取能量,地球又为何一定荒芜。

不能只用物理定律去观察和分析宇宙的发生与发展,应该将物理定律、生命定律和信息定律结合起来看宇宙的发生和发展。更要将三个空间的定律结合起来,改变这个宇宙甚至泛宇宙的发展。

